

# A corpus investigation of formulaicity and hybridity in legal language a case of EU case law texts

Trklja, Aleksandar

*License:*

Other (please specify with Rights Statement)

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Trklja, A 2018, A corpus investigation of formulaicity and hybridity in legal language a case of EU case law texts. in *Phraseology in Legal and Institutional Settings: A Corpus-based Interdisciplinary Perspective*. Law, Language and Communication, Routledge, pp. 89-106.

[Link to publication on Research at Birmingham portal](#)

**Publisher Rights Statement:**

This is an Accepted Manuscript of a book chapter published by Routledge in *Phraseology in Legal and Institutional Settings: A Corpus-based Interdisciplinary Perspective*, on 7/8/2017, available online: <https://www.routledge.com/Phraseology-in-Legal-and-Institutional-Settings-A-Corpus-based-Interdisciplinary/Goetz-Roszkowski-Pontrandolfo/p/book/9781138214361>

**General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

**Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

# **A corpus investigation of formulaicity and hybridity in legal language: a case of EU case law texts**

*Aleksandar Trklja*

## **1. Introduction**

The aims of the present paper are twofold. First, it contributes to the field of legal linguistics by providing evidence for the use of formulaic and hybrid expressions in legal language. The study will in particular focus on judgments of the Court of Justice of the European Union (CJEU) Second, it proposes new empirical methods for the study of discourse organization on the one hand and of semantic and grammatical profiles of lexical items on the other.

Traditionally, legal linguistic studies focus on the recurrent use of legal terms that have specific ideational meanings (e.g. Tiersma, 1999) or on the impact that the rigid nature of formulaic expressions might have on law. However, there are no theoretical or methodological reasons why the study of repetition in legal language should be restricted to legal terminology understood in a narrow sense. In fact, legal terminology is part of formulaicity as a more general phenomenon. Apart from the fact that it is often very difficult to distinguish between legal and non-legal meaning of lexical items (e.g. Goźdz-Roszkowski, 2011) formulaicity includes the types of expressions that have non-ideational meaning. Montolio (2001) and Goźdz-Roszkowski, (2011), for example, illustrate how recurrent expressions contribute to the textuality of legal texts. McAuliffe (2009) also shows that drafters of the judgments at the Court of Justice of the European Union (CJEU) are constrained by the formulaic style of these documents. This paper demonstrates that the investigation of types of formulaic expressions that signal discourse organization is key for an understanding how information and argumentation develops in legal texts.

Another well-known feature of legal language is the use of idiosyncratic expressions. The expressions typically discussed in the literature (e.g. Charrow, Crandall and Charrow, 1982; Tiersma, 1999), include legal archaisms (e.g. *further affiant sayeth not, be it known*) or formal and ritualistic words and phrases (e.g. *Wherefore the Plaintiff prays for relief as follows*). However, more recent studies (Kermas, 2010; McAuliffe, 2011; Biel, 2014) indicate that such expressions in legal texts can also be created through translation. This is especially the case in the context of EU institutions where communication takes place to a large extent through translation. The language of EU institutions is described as being ‘strange’ because it departs from ‘normal’ use observed in non-translated texts (Born, 1995; Muhr, 2002; Tirkkonen-Condit, 2001). This phenomenon is

referred to as *hybrid language* (Schaeffner and Adab, 2001; McAuliffe, 2011). However, linguistic aspects of hybrid languages have not been investigated in a systematic manner. The present paper demonstrates that this gap can be filled by using an approach based on a quantitative comparative analysis of local grammars. Hybrid expressions are considered as lexical items which are produced through translation into a target language, and the semantics of which depart from the semantics observed in ‘standard’ use.

Section 2 sets out the notions of formulaicity and hybridity. The subsequent two sections describe the role of the CJEU, as well as the data, methodology and theory used. Methods of analysis and results are presented in Section 5 in the form of three individual studies.

## **2. Formulaicity and Hybridity**

In one of the earliest linguistic investigations of formulaic language, Pawley and Syder (1983) suggest that language users’ mental lexicon consists of holistically stored linguistic sequences. They refer to these sequences as ‘lexicalized or institutionalized sentence stems’. These units are of clause length or longer and according to the authors such expressions facilitate language processing. A similar view was expressed by Sinclair (1991) in his formulation of ‘idiom principle’:

“The principle of idiom is that a language user has available to him or her a large number of semi-preconstructed phrases that constitute single choices, even though they might appear to be analysable into segments” (Sinclair 1991: 110).

Subsequent empirical research both in corpus linguistics (e.g. Biber, 2009) and psycholinguistics (e.g. Schmitt, 2004; Wray, 2005; Conklin and Schmitt, 2012) provided further evidence for these claims.

Biber and Conrad (1999) proposed a corpus-driven method of investigation of formulaic language. The method, which was further elaborated in subsequent studies (e.g. Biber et al, 2004; Biber, 2009) focuses on the distribution of frequently recurring fixed sequences of words called *lexical bundles*. Lexical bundles can be of various lengths but are typically 3-to 6-word long sequences. These sequences are incomplete structural units both in semantic and grammatical terms and can be classified into different classes according to their functions. These functions include epistemic meaning, the expression of attitudes, indication of references and signalling textual or discourse organization. To date the most comprehensive investigation of the lexical bundles that have textual

function is Nesi and Basturkmen (2006). Gozdz-Roszkowski (2011) identifies typical lexical bundles in U.S. legal texts and demonstrates how their distribution reveals variation of legal genres in American legal English.

The present study adopts the general principle of the analysis of lexical bundles, however that principle is further developed here by introducing two new methodological features. First, previous studies say almost nothing about the degree of formulaicity of texts. They are concerned with the distribution of lexical bundles across registers and ignore individual texts. This paper addresses that issue using a new approach, set out in Section 5 below. Secondly, as mentioned above, lexical bundles are structurally incomplete units. In contrast, in the present paper focuses on units which are not only functionally and structurally complete but which are also associated with a specific textual position. These units are part of information structure in language and they signal discourse organization of texts.

The paper examines the following research questions in relation to formulaicity:

1. To what extent are CJEU judgments formulaic?
2. How do formulaic units contribute to the discourse organization of CJEU judgments?

The notion of hybridity that has its origin in the 19<sup>th</sup> century discourse of race (Young, 2000) was first introduced to social science as an analytical tool by Bakhtin (1981) in *The Dialogic Imagination*. Since then, the term has been used in various disciplines such as social anthropology (Hannerz, 1987), cultural studies (Bhabha, 1994) or translation studies (Schäffner and Adab, 2001). What is common to different definitions offered so far is that hybridity is seen as a force that creates new cultural forms, undermines the established ways of thinking and increases variety.

Schäffner and Adab argue that hybridity is the defining feature of translated texts because these texts do “not conform to established norms and conventions” (Schäffner and Adab, 2001 169). The ‘strangeness’ of translated texts relies on the fact that they contain linguistic features that do not occur in non-translated texts. It is because of these features that the language of translated text deviates from the use of language in non-translated texts (Bond, 2001). Neubert (2001) warns that although translated texts might contain hybrid texts it would be wrong to consider them in their totality as hybrid. The present paper adopts the latter position.

In the context of legal studies, McAuliffe (2011; 2013) argues that legal judgments produced by the CJEU are hybrids because a) they are produced in a multilingual context and b) they are produced

through translation. In an in-depth description of the work and procedures employed by the Court the author demonstrates how various types of legal texts are first produced in French and then translated into other languages. French serves as the working language of the CJEU but the majority of those involved in the text production process are non-native French speakers. In addition, these drafters work under time pressure and are expected to ensure coherence of EU law. All these factors contribute to the stylistic peculiarity of CJEU judgments.

The above research serves as a starting point to address the third research question.

### 3. Can we find linguistic evidence of hybridity of CJEU judgments?

#### 3. CJEU Judgments

The Court of Justice of the European Union is the highest court in the EU legal order. The main role of the CJEU as stated in Article 220 Treaty of Rome is to “ensure that in the interpretation and application of the Treaty the law is observed”. The Court delivers binding judgments regarding questions of interpretation of EU law in up to 24 language versions (the 24 official languages of the EU) and those judgments constitute an EU case law.<sup>1</sup> Although the Court produces judgments in all EU official languages, for practical reasons that institution works in one language – French. Thus, all judgments are first drafted in French and then translated into the other EU official languages. However, only one version of a judgment is considered ‘official’, the version in the language of the case, which is rarely French.<sup>2</sup> In other words the official version of a judgment is more often than not a translation. A judgment is a collegiate text, the (French language) final version of which is agreed on by the relevant judges in a chambers secret deliberations. This final, French language, version is translated into the other official EU languages by lawyer-linguists who are professional lawyers by vocation but who are usually not trained translators. It is not uncommon that lawyer linguist at one point moves to the position of référendaire who are legal assistants in judges’ cabinets (McAuliffe, 2011). CJEU judgments are thus multi-authored texts, created through translation. From the fact that the authentic version of a CJEU judgment is usually a translation it can be concluded the Court perceives translation as a neutral medium of communication that does not have any important impact on the form and content of EU case law.

---

<sup>1</sup> At the time of going to press there are 24 official EU languages. 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.

<sup>2</sup> Each case has a ‘language of procedure’ and only the judgments produced in the language of procedure are considered ‘authentic’, in spite of the fact that they are usually translations. For more information on the language regime at the CJEU see McAuliffe 2011, 2013.

#### 4. The corpus of CJEU judgments

CJEU judgments have been collected in the present study and stored into several corpora and sub-corpora<sup>3</sup> to answer the three research questions introduced in the previous section.

The degree of formulaicity in CJEU judgments is investigated in a comparative study using two sets of sub-corpora. The first set consists of the English, French, German and Italian language version of CJEU judgments. Because an analysis of all available judgments in all official languages would be prohibitively time consuming for the purposes of the present paper, the relevant sample is restricted to 1140 CJEU judgments and to four languages. These judgments were produced in the period between 1955 and 2011 and the corpus comprises between 8 and 9 million tokens depending on individual languages. These judgments make up the EU *acquis communautaire* case law and are as such considered to be the most important judgments in EU law. The second set contains judgments produced by constitutional or supreme courts of EU member states in which English, French, German and Italian are official languages. This second corpus covers the same time period as the corpus of CJEU judgments and contains approximately the same number of words. Since the former three languages are used in two member states this second set includes judgments from seven national courts<sup>4</sup>. National constitutional or supreme courts were chosen since they are the closest types of national courts for comparative purposes to the CJEU.

Owing to limits of time and requirements regarding length of the present paper, the second and the third questions are addressed only by focusing on the English language versions of CJEU judgments. Analysis of the use of lexical items signal discourse organization of judgments is based on CJEU judgments only. The occurrence of hybrid expressions is studied in a comparative analysis between CJEU judgments, UK Supreme Court judgments and texts from the British National Corpus<sup>5</sup>.

In addition to courts' decisions, the texts of judgments contain other sections consisting of summaries of facts and/or law, keywords or party names. Since the present analysis focuses on the language of decisions such sections have been removed from corpora by means of a Python script

---

<sup>3</sup> These corpora have been compiled within the European Research Council (ERC) project 'Law and Language at the European Court of Justice.'

<sup>4</sup> These courts are: *Verfassungsgerichtshof Österreich* (the Constitutional Court of Austria), *Bundesverfassungsgericht* (German Federal Constitutional Court), *UK Supreme Court*, the *Supreme Court of Ireland*, *Corte costituzionale della Repubblica Italiana* (the Constitutional Court of the Italian Republic), *Conseil constitutionnel* (French Constitutional Council) and *Grondwettelijk Hof* or *Cour constitutionnelle* (the Constitutional Court of Belgium).

<sup>5</sup> A fuller and more cohesive study relating to these questions is being carried out in the ERC-funded 'Law and Language at the European Court of Justice' project. For further information see [www.llecj.karenmcauliffe.com](http://www.llecj.karenmcauliffe.com)

created for the purposes of the present paper. Other tools used in the studies and the procedure of analysis are described in Section 5.

## **5. Theoretical and methodological issues**

### **4.1 The theory of information distribution in text**

Although sentences unfold in a linear fashion in a text, discourse has a structure (Hobbs, 1985). Texts consist of discourse units which are semantically organized in terms of various types of relations. Discourse units have different discourse values or communication functions which are “determined largely with respect to the interaction between sentence meaning and context” (Crombie, 1985: 2). This means that information development in discourse can be understood by looking at the ways in which sentences are related to each other.

The basic functional units in information structure are Theme and Rheme (Halliday, 1985) that give a clause the character of a message. These two units are associated with specific positions in clause. Theme refers to all the elements in a clause that start from a clause boundary and end with a finite verb and Rheme covers the rest of the clause. Theme serves as “the point of departure for the message ... that with which the clause is concerned” (Halliday, 1985: 38). The content of the message is developed in Rheme which is typically associated with new information. The elements that occur in the Theme position, therefore, signal how the message will develop and the content of this message is located in Rheme. According to Halliday, there are three kinds of Themes: ideational or topical, interpersonal and textual. Ideational Themes indicate the propositional content of a clause or message. Interpersonal Themes signal the writer-reader relationship and Textual Themes are about how the distribution of information is signalled.

Building on this theory of information structure, Fries (1981) proposes a ‘method of development’ that goes beyond the analysis of clause relations and that demonstrates how information flows at the level of text. Lexical items that occur in the Theme position serve as cohesive ties and the method demonstrates how ideas develop in texts. Fries’ method, in other words, implies that thematic items do not only signal information structure of a clause and sentence but also of a discourse. For example, ‘In those circumstances’ typically indicates that the information in a given clause serves as a conclusion that follows from a unit of information provided in the previous stretch of discourse. This is illustrated in the following extract. Sentences are enumerated for the ease of referencing.

1. 1) The Italian Government further claims that, without a guaranteed market outlet, the cultivation of durum wheat would disappear from the regions of the Mezzogiorno where it is practised. ... 2) The statistics supplied to the Court show a steady increase in the market share held by pasta products made exclusively from durum wheat in other Member States in which they already face competition from pasta made from common wheat or from a mixture of common wheat and durum wheat. 3) **In those circumstances**, it is clear that the fears expressed by the Italian Government as to the disappearance of durum-wheat growing are unfounded.

The first sentence introduces a claim for which contradictory evidence was provided in the second sentence. The textual Theme ‘In those circumstances’ from the third sentence finally refers to the content of the previous sentence and signals that if this is true then it can be concluded that the original claim is wrong and should be rejected.

The studies conducted to date (e.g. Halliday, 1985; Fries, 1995; Martin, 1995) have been concerned with the flow of information investigated in terms of ideational meaning or ideational motifs. It means that the contribution of interpersonal and textual Themes has been ignored. Since the objective of the present paper is to investigate how the flow or organization of information in text is signalled, the most relevant kinds of Theme are those that denote textual meaning. In addition, unlike previous studies which are concerned with short texts, the results presented in the study below derive from a quantitative analysis.

Halliday (1985) proposes a system of logico-semantic relations that accounts for relations between clauses<sup>6</sup>. This system provides a sound basis for the study of the functions of textual Themes. Due to word limits this system is only briefly described here and for a more detailed explanation an interested reader is referred to Halliday (1985) or Martin (1992). Halliday distinguishes between three kinds of logico-semantic relations: Elaboration, Extension and Enhancement.

Elaboration items serve to signal that one clause “elaborates on the meaning of another by further specifying or describing it” (Halliday, 1985: 203). These items indicate that the subsequent clause does not contain new information but instead provides further characterization of a previous clause. There are three types of Elaboration items:

---

<sup>6</sup> Although called *logico-semantic relations* these are purely semantic relations because they do not include logically valid inference relations between propositions. However, for the sake of clarity the established term is used here.



- Exposition indicates restatement (e.g. *in other words*);
- Exemplification indicates providing examples (e.g. *for example*);
- Clarification indicates further clarification of a message (e.g. *in other words*).

Extension items signal that a clause adds new information to a previously introduced message.

Distinctions can be made between:

- Addition indicates adding new content to an existing message (e.g. *in addition*);
- Alternation indicates variation in the content of a message (e.g. *in the alternative*);
- Variation indicates replacement of the content of a previous message (e.g. *on the contrary*).

Enhancement items signal qualifying the content of a message by reference of time, place, manner, cause or condition.

- Temporal relations indicate successive or simultaneous order (e.g. *at the same time*);
- Spatial relations indicate at what point something happened (e.g. *in the present case*);
- Manner relations indicate by what means something happened (e.g. *in this way*);
- Causal conditional indicate for what purpose something happened (e.g. *for this reason*).

Information structure and logico-semantic relations are defined in terms of the position of lexical items in a clause. Since the automatic identification of textual Themes in CJEU judgments at clause level is not possible and a manual analysis is not practically possible for the purposes of the present paper, the present study is restricted to the sentence level.

Following Fries' (1981: 135) findings that "the information contained within the Themes of all the sentences of a paragraph creates the method of development of that paragraph", it is assumed that the study of textual Themes in sentences can indicate the organization of discourse at paragraph level. In addition, Moore (2016: 10) argues that "the fundamental function of INFORMATION STRUCTURE is to divide the flow of discourse into manageable units ... that punctuation functions to divide written discourse into manageable units". Relying on this, it is assumed that commas demarcate the items that have textual meaning in text. The units of analysis selected to deal with the second research question are all multi-word sentence initial expressions that end with a comma.

#### **4.2 The local grammar approach**

The local grammar approach was developed by Gross (Gross, 1987, 1993, 1997) and its purpose was to account for how rules locally constrain co-occurrence of words. The approach relies on

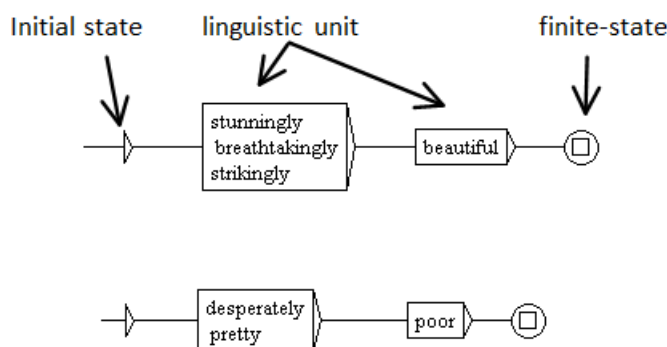
Harris' distributional theory of language (e.g. Harris, 1954, 1988) and the theory of finite-state local automata (e.g. Roche and Schabes, 1997).

First it is assumed that “the occurrence of each word in an utterance depends on the occurrence there of an element—any element—of some stated subset of words...” (Harris, 2002: 216). This claim is similar to the notion of s(semantic)-selection introduced by Chomsky (1965) which specifies restrictions between lexical items that co-occur in the same textual context. With any lexical item there will be a limited number of co-occurring items that will constitute a sub-set within a general grammar category. For example, *beautiful* and *poor* are both adjectives that can be preceded by adverbs. However, it does not mean that these two adjectives select any type of adverbs. Thus, according to the BNC *beautiful* collocates with *stunningly*, *breathtakingly*, *strikingly* but not with *desperately* or *pretty* which are found with *poor*.

The co-selection of lexical items<sup>7</sup> is of a finite-state nature because “short range constraints between words in sentences are crudely accounted for by Markovian [or finite-state] model” (Gross, 1997: 330). In other words, local co-selection relations between lexical items unlike relations in general syntax include restricted options because they involve constraints of combination of words. Finite-state automata are powerful devices that can account for constraints operating on a local syntactic level (Roche and Schabes, 1997). These constraints allow or preclude particular classes of combinations (Harris, 1991). One simple example of a finite-state automaton is illustrated in the following figure which presents the finite-state nature of the co-occurrence of *beautiful* and *poor* and their collocates. As can be seen, each finite-state automaton has one initial state and one finite state denoted by the leftmost arrow and the rightmost square respectively. These states simply mean that a linguistic unit has a beginning and end. The central circle represents an inventory of options available for the construction of a lexical item. As in this example, the options include sets or paradigms that might contain one or more elements.

---

<sup>7</sup> Chomsky (1957) argues that finite-state grammars present a model which is too simple to describe the syntax of natural languages. They have a too limited expressive power to capture complex combinatorial options available in the syntax of general language.



**Figure 1: An example of a finite-state automaton**

The diagram represents only a segment of the local grammars of *beautiful* and *poor*. In reality the number of elements would be much higher and the relations between them much more complex.

The structure of strings generated through a local grammar can be represented by means of phrase structure rules introduced by Chomsky (1957). This type of string rewriting system has an initial string and a string derived by means of a rule. Although this algorithmic device has been mainly used to show relations between lexical items and general parts-of-speech categories in the third study in Section 5 it is demonstrated that it can be adopted to describe relations at the local grammar level.

Staying with the same example, the adverbs observed above with *beautiful* do not occur with equal likelihood (*strikingly* is found in a stronger collocation in this context than *stunningly* which is found in a stronger collocation than *breathtakingly*). This is because linguistic systems are probabilistic (Halliday, 1991: 42) and with every lexical item there will be inequalities or grading from most to least likely collocates (Harris, 1991). Through the investigation of these inequalities we can identify typicality of co-occurrence of lexical items.

It is the probabilistic nature of co-occurrence of lexical items that can help to distinguish between hybrid and non-hybrid expressions. In other words, the question whether a lexical item ‘departs’ or not from a usual usage will depend on that whether the likelihood of co-occurrence of collocates corresponds or not to that found in a corpus that represents a standard language use. The finite-state automata make it possible to capture the types of lexical items that characterize the use of hybrid items in a systematic manner.

## **6. Linguistic analysis of formulaicity and hybridity in CJEU judgments**

### **6.1 Study I: Degree of formulaicity**

This study addresses the first research question introduced in Section 2. In order to deal with this question, it is first necessary to establish a method of measuring the degree of formulaicity of CJEU judgments and to find a model of investigating this extent at the textual level. To deal with the first requirement the study compares the degree of formulaicity in CJEU and in national judgments. The judgments produced by supreme or constitutional courts of EU member states will serve as yardsticks that reflect ‘standard’ level of formulaicity in the register of legal judgments. The extent of formulaicity in CJEU judgments will therefore be measured with respect to these standard values. The second requirement is captured by calculating the percentage of repeated expressions in individual judgments.

The units of analysis used in the study are all repetitive lexical bundles that are at least 5-words long. The only two criteria for deciding the length of lexical bundles are the size of corpora and the frequency of lexical items (Biber, 2006). The preliminary investigation shows that 5-word lexical bundles are sufficient to mirror adequately the occurrence of formulaic expressions in the relevant corpora. Owing to the practice of citation and to the principle of precedent in common law systems in particular (e.g. Brenner, 1992), judgments might occasionally contain longer textual chunks. For this reason, those expressions which are longer than 5-word lexical bundles are also included in the present analysis.

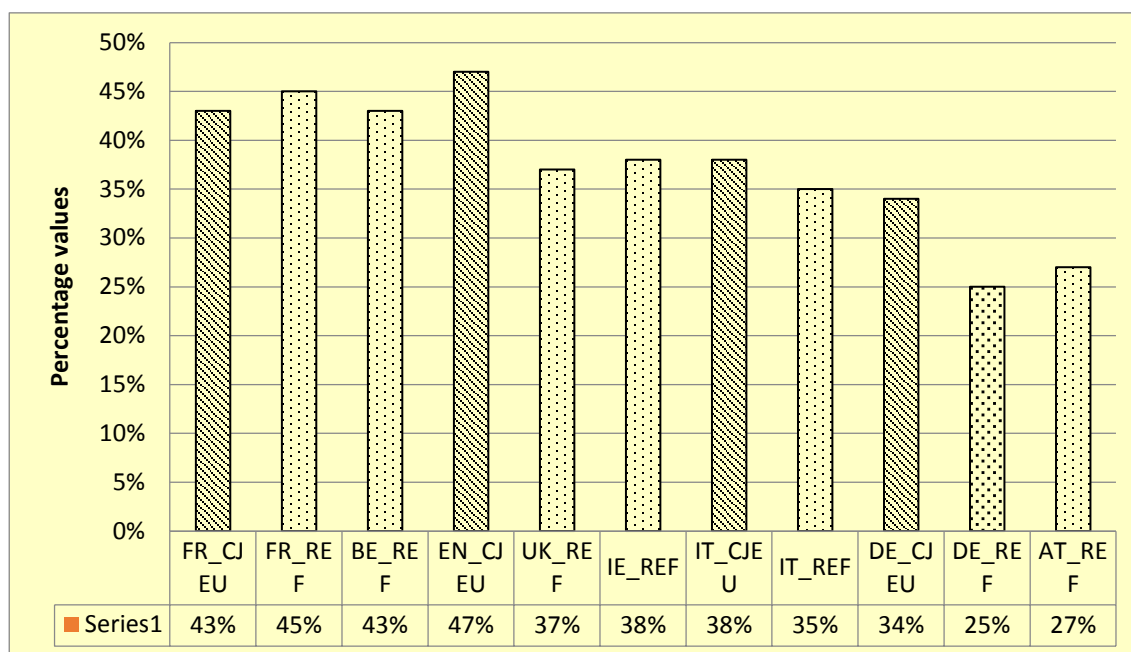
Preliminary analysis indicates that there are two factors that can influence results: a) the number of texts compared and b) the length of texts. In addition, corpora with fewer texts tend to have a lesser degree of repetition and corpora with a larger number of texts have more variations in the length of texts. For example, 50 repeated words found in a text that has 300 words comprises 17% and in a text that has 3000 words less than one percent. To overcome these problems a range of samples that consist of 100 texts (from CJEU and reference corpora) was created. In both corpora the number of texts from different years varies and there are more texts from more recent periods. In the CJEU corpus 32% of the texts analysed are from the 1990s, 23% from the 1980s, 22% from the period between 2000 and 2010, 4% from 1960, about 1% from the period between 2010 and 2012 and less than one per cent from the 1950s. For this reason, texts are selected at the proportional rate for each decade. Whenever possible the same proportion and decades were reflected in the reference corpora. There are also significant differences in the style of reporting judgements between different countries therefore an attempt was made to select texts of similar size and to exclude those that are either very short or very long.

Because investigations of this kind are extremely time and resource-consuming the findings presented below are based on analysis of five samples of 100 judgments. For the same reason, it

was not possible to carry out an analysis covering all 24 official languages. Instead, the study is restricted to four languages: English, German, French and Italian language. The former three languages are used in two EU member states and it means that the study covers four languages and nine national courts.

At the next stage, a calculation was carried out to show in terms of percentage values expressions that occur in other texts. The analysis was divided into two steps. First, all 5-word long or longer multi-word expressions that occur in samples and in the rest of the corpus are identified. This analysis was carried out in both sets of corpora. A Python script was created to compare each text from the five samples to all other texts from the corpus in order to identify repetitive multi-word expressions that occur in study texts. After that the average values for these texts were calculated and results between CJEU and national judgments were compared.

The average values of degree of formulaicity for all relevant judgments are displayed in Table 1. Bars with striped lines denote the results of CJEU judgments and dotted bars of national judgments.



**Table 1: Degrees of formulaicity in CJEU and national judgments**

It can be observed that with the exception of judgments of the French Constitutional Court in all other instance CJEU judgments tend to contain more formulaic expressions than national judgments. The first conclusion to be drawn is that there is more similarity in terms of formulaicity between national judgments than between them and CJEU judgments. The high formulaicity degree of CJEU judgments and this difference highlight the unique linguistic style of these judgments.

Differences that can be observed across languages are because of structural differences between languages which have an impact on the size and number of n-gram constructions. These differences, therefore, do not demonstrate that, for example, English judgments are more formulaic than German judgments. German is a synthetic and English is an analytic language which means that the same unit of meaning can be realized in the former as one word and in the latter in two or more words<sup>8</sup>.

A test of significance was conducted to determine whether the differences between formulaicity values in CJEU and national judgments are statistically significant. A one-way Anova test was performed for German, English and French because these data sets contain three variables and an independent t-test for Italian because there are only two variables here. The results indicate that the difference between CJEU and national judgments is statistically significant ( $F=12.8$ ,  $p=.000$ ) only for English versions of judgments. Statistically significant differences cannot be observed in French ( $F=1.5$ ,  $p=.218$ ), German ( $F=1.8$ ,  $p=.156$ ) and Italian ( $t\text{-score}(298)=-1.3$ ,  $p=.200$ ).

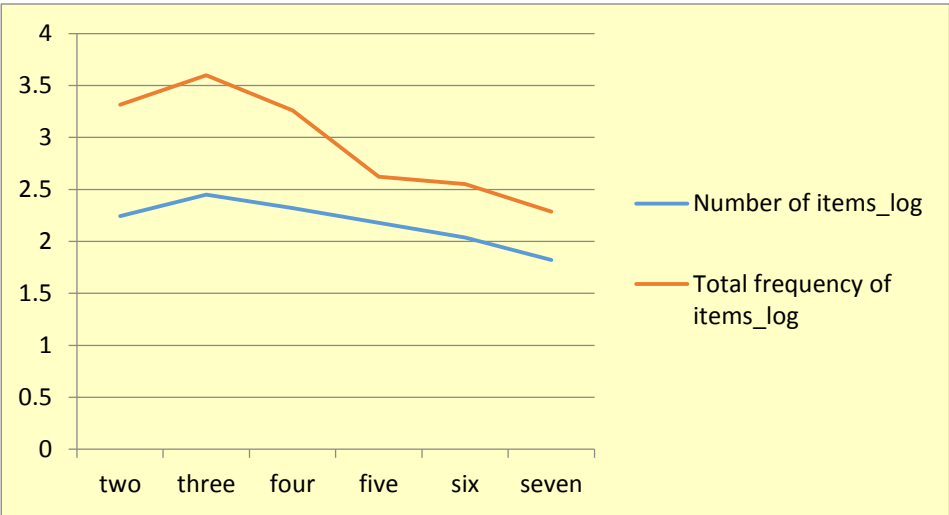
The results demonstrate that the method employed can successfully provide the values of the degree of formulaicity at the textual level. These values indicate that formulaicity is one of the features of CJEU judgments because the degree of formulaicity tends to be beyond the standard level observed in national judgments. The statistical tests indicates that formulaicity is especially strongly associated with the English version of CJEU judgments. Finally, the fact that judgments of the French Constitutional Court have a high degree of formulaicity is indicative. French is the working language of the CJEU and French administrative law served as a model for EU case law in its formative years. From this once can assume that the legal French has influenced the linguistic shape of CJEU judgments. This hypothesis deserves further investigation.

## **6.2 Study II: Discourse organization and formulaicity**

As mentioned above in Section 2, the units of analysis selected to address the second research question are all sentence initial multiword expressions that are at least 2-words long and that end with a comma. In the first stage of analysis, 1760 linguistic units that meet the above criteria were extracted in the English version of CJEU judgments by means of CQP tools. These units are between two and six words long. Their frequency and quantity are displayed in Table 2. Since the figures for the two variables differ in scale the results are summarized in terms of their log values.

---

<sup>8</sup> To get more comparable results of formulaicity across languages it would be necessary to identify appropriate lengths of lexical bundles for each language. Although of interest, this is beyond the scope of the present paper.

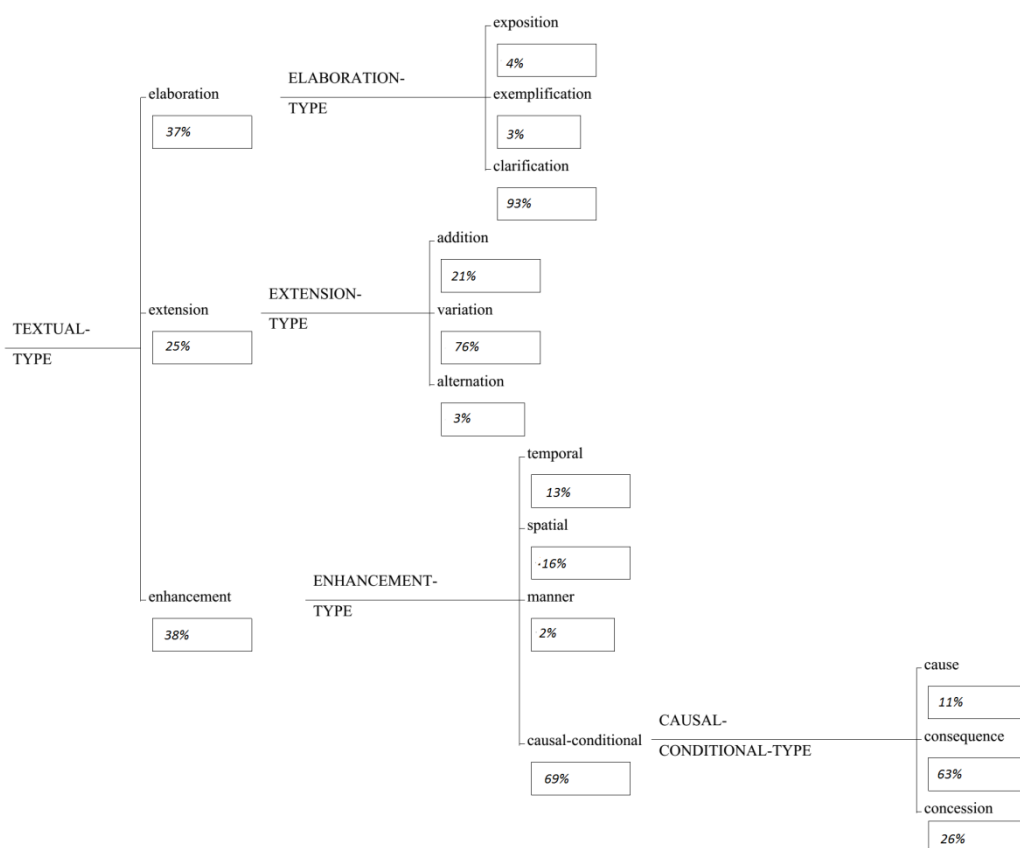


**Table 2: The frequency and number of Themes in CJEU judgments**

Biber (1995) reports that correlation between the length and frequency of lexical bundles can be observed in his data. In contrast, in the present data the most numerous items are not the shortest linguistic units. This might suggest that structurally and functionally complete formulaic expressions have preferences regarding length. Table 2 also shows that the number of items corresponds to their frequency, which means that once their length is established the frequency of lexical items can be predicted.

In the next stage of analysis, the identified linguistic units were classified into three types of textual Themes following Halliday’s system. First, it is assumed that the least frequent items do not contribute to the formulaicity of judgments. This assumption is justified by the results which show that items occurring five times, or more frequently, make up 67% of the frequency of all items identified. These more frequent items, therefore, reflect the typical use of thematic items and they were thus selected for further investigation. Out of 248 Themes 49% have a textual function, 26% interpersonal and 25% ideational function. In terms of frequency of occurrence, 80% are textual Themes and the other two types 10% each. These results indicate that the beginning of sentences in CJEU judgments typically serve to signal organization of discourse. It can also be concluded that the same items tend to be more often re-used when it denotes the meaning of textual than interpersonal or ideational Themes.

At the next stage all textual Themes identified (108 items) are classified into categories in terms of the system of logico-semantic relations. Since the focus of the study is on the method of development of texts, all ideational and interactional Themes are excluded from further consideration. The following graph displays the distribution of textual Themes in relation to all categories and sub-categories from the system of logico-semantic relations.



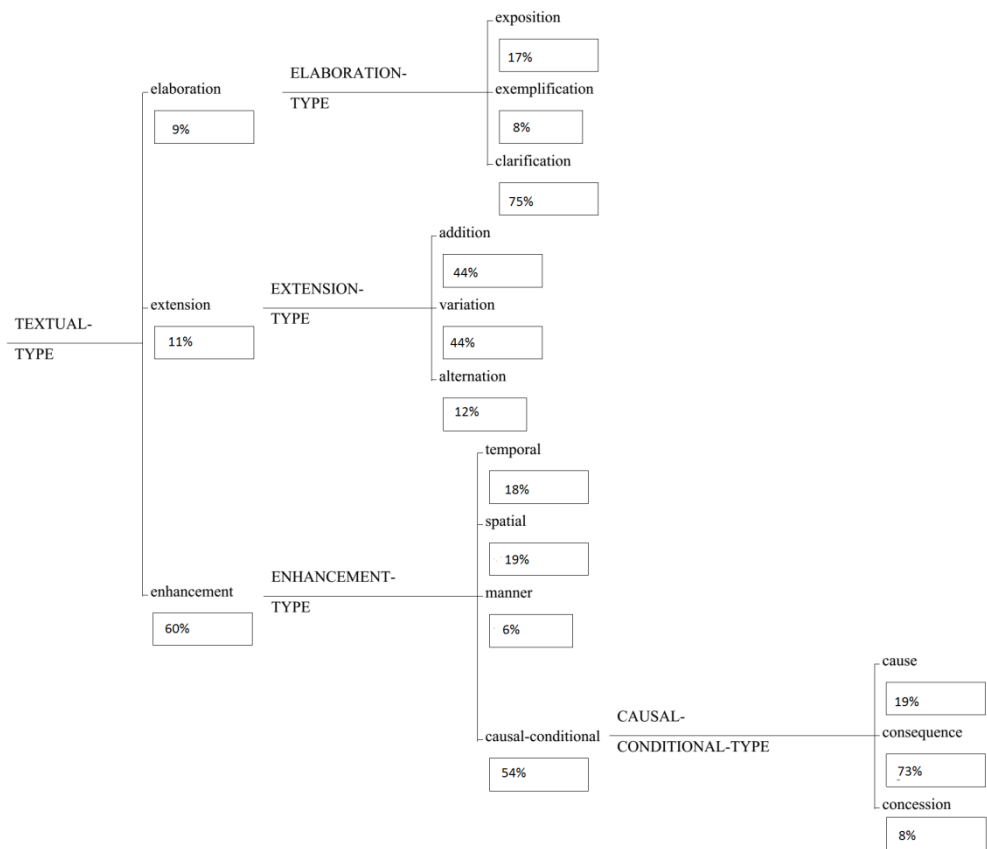
**Figure 2: Frequency of textual Themes in terms of logico-semantic relations**

There are, at first sight, no important differences between three kinds of relations (Enhancement, 38%; Elaboration, 37%; and Extension, 25%) but greater variations can be observed with respect to more delicate options. Thus, textual Themes that denote Causal-conditional relations occur with higher likelihood than other types of Enhancement. It follows that it is very typical for the Court to reason its decision by first developing certain points and then clarifying its position towards issues expressed by means of these points. This type of relation is most frequently realized by means of the lexical items such as *In those circumstance*, *On those grounds*, *In that case*, *For that reason*, *In such circumstances*, *As a result*, *It follows that*, *On that basis*, *That being the case*. Similarly, Clarification is the most typical kind of Elaboration and Variation is the most typical kind of Extension found in CJEU judgments. This means that textual Themes in CJEU judgments often signal that a subsequent piece of discourse will contain an additional explanation or correction or contrasting views. The most frequently used Clarification items are *In particular*, *On the one hand*, *In effect*, *In this connection*, *In any case*, *What is more*, *In reality*, *In essence* and the most frequently occurring Verification items are, *On the contrary*, *On the one hand*, *By contrast*, *In contrast*.



Grammatically, 92% of all textual Themes are prepositional phrases. The items from the same categories usually consist of identical grammatical and lexical elements. For example, the majority of Clarification items have the structure <in + DET + connection|regard|respect> such as in *In that regard*, *In that respect* or *In this connection*. Vertical bars here indicate alternative options and DTE denote determiners. To give another example, Condition items have the following structure: <In|Under + those|these|those|the + circumstances|situation>. These results indicate formulaic nature of textual Themes by showing that individual types of textual Themes are made up of restricted sets of lexical items.

The graph above displays the distribution of textual Themes in terms of their frequency. The following graph, on the other hand, shows the number of items found within individual categories and sub-categories. The items belonging to the category Enhancement appear to be most numerous. However, this has to do with the nature of taxonomy rather than with linguistic devices used in CJEU judgments, because the Enhancement category contains more sub-categories than the other two categories. One striking feature in the data is that two types of the most frequently used types of textual Themes (Clarification and Variation) have lower figures in this second graph. Thus, the Clarification items that in terms of frequency make up 93% of all Elaboration items occur with the value 75% in terms of the number of items per categories. The respective values for the Variation items are 76% and 44%. At first sight, this does not seem to be the case with the items from the category Consequence but this is only true as long as we compare the three most delicate sub-categories of Enhancement items. However, if we compare figures globally we can see that the value for this category is 34% in terms of frequency and 23% in terms of the number of items. All these differences demonstrate that the most frequently used types of textual Themes tend to be re-used more often than the less frequent types. It follows that drafters of CJEU judgments tend to re-select from a small set of available resources. This has a direct impact on how the flow of information and reasoning is organized in CJEU judgments.



**Figure 3: Numbers of textual Themes in terms of logico-semantic relations**

The results of the present analysis show that textual Themes serve as formulaic expressions that signal discourse organization of CJEU judgments. The flow of information in these judgments is typically based on the devices signalling that the subsequent discourse provides more information, contrasting views or that the content of the subsequent discourse is conditioned by what was said before. In terms of argumentation theory it can be said that the Court puts emphasize on providing clear arguments and creating logically valid inferences.

### 6.3 Study III: Hybrid expressions

To identify potential candidates to be considered as hybrid expressions, two keyword analyses were carried out using WordSmith tools (Scott, 2012). As explained in Section 3 above, the reference corpora used in the study are the BNC and a corpus of judgments of UK Supreme Court (UKSC). The purpose of conducting two instead of one analysis is to identify expressions with high keyness values with respect to general English and the UK legal English. These values indicate that items depart from the standard use as reflected in the BNC and the UKSC.

First, the keyword lists obtained in the two analyses were compared and only those items that occurred with high keyness value in both analyses were selected for further analysis. Since the purpose of the present study is to explore the occurrence of hybrid expressions created through translation, only those linguistic items with suffixes historically borrowed from French were taken into account. The assumption here is that the influence of French as a source language will be more visible for those English lexical items which are historically related to French. Some of the suffixes imported from French in the Middle English period which are still in use are *-bility*, *-ble*, *-nym*, *-ary*, *-ment*, (Dalton-Puffer, 1996; Zergollem-Miletic, 1997). These lexical items are mainly nouns and the final stage of analysis involved the creation of a keyword list of all nouns that occur in CJEU judgments. The candidates identified in the data in this way include *interoperability* (198), *compatibility* (56), *competition* (5), *concentration* (17), *consumption* (38), *cooperation* (24), *distribution* (36), *inadmissibility* (88), *notification* (46), *objections* (26), *production* (19), *treatment*(92). The numbers in parenthesis indicate words' rank in the keyword list of nouns that contain 8300 nouns that occur at least five times in the corpus of CJEU judgments.

To explore whether these items depart from established use it is necessary to fully understand their distribution. Such analysis is demonstrated here by exploring the use of the lexical item *compatibility*. First its most typical collocates are identified across three corpora and the results are compared. Then the local grammar approach is used to conduct a fine-grained analysis of its grammatical and semantic profile.

The lexical item *compatibility* is 22 times more likely to occur in CJEU judgments than in the BNC and 11 times more likely in CJEU judgments than in the UKSC. It follows that it generally prefers occurring in the legal register. It occurs in a verb and noun phrase; i) as an argument in the frame V +*compatibility* ii) and as a complement in the frame N + of + *compatibility*.

With regard to the frequency of occurrence of verbs and nouns from these two grammatical frames, the items that occur only once can be ignored because they do not reflect typical collocates. The lexical item *compatibility* occurs with a larger number of verbs in CJEU judgments (37) than in the BNC (23) or UKSC (29). Also, there are four nouns that colligate with *compatibility* in CJEU judgments, one in the UKSC and none in the BNC. The BNC is 17 and UKSC 5 times bigger than the corpus of CJEU judgments and one would therefore expect inverse results.

A closer investigation of the types of words that occur with *compatibility* shows that it is most typically associated with expressions that denote assessment. There are eight verbs with this meaning in the first structure and four nouns in the second structure. As can be observed from Table

3, all these items occur with the highest likelihood in CJEU judgments. The figures show raw frequency values and the occurrence of lexical items per million words. For example, <assess> is 84 and <examine> 53 times more likely in CJEU judgments than in UK judgments. The first item is 412 times more common in the CJEU corpus than in the BNC. Using the understanding of hybridity formulated in Section 2 it can be concluded that *compatibility* occurs in hybrid expressions in CJEU judgments.

Lexical items	Freq_CJEU	Per Mil_CJEU	Freq_UK	Per Mil_UK	Freq_UK	Per Mil_UK
assess	49	7.332098	3	0.087351	2	0.017811
examine	41	6.135021	4	0.116467		
consider	10	1.496347	10	0.291169		
review	4	0.598539	2	0.058234		
prejudge	2	0.299269				
enquire	2	0.299269				
appraise	2	0.299269				
verify	2	0.299269				
assessment	16	2.394154				
examination	15	2.24452				
analysis	4	0.598539	2	0.058234		

Table 3: *Collocates of compatibility*

However, the collocation analysis is insufficient because it does not provide a comprehensive description of the distribution of this item. It is only through such a description that we can understand the ‘strangeness’ of a hybrid expression. At the next stage of analysis the rules that govern the relationship between *compatibility* and its collocates are established.

Such a description is presented below first in the form of rewriting rules (Chomsky, 1957) and then a finite-state diagram. [2] and [3] below show first the structure of the verbal and nominal phrase in which *compatibility* occurs. After that, the members of the parts-of-speech categories are specified. The local grammar sets identified here are ASSESS and COMPATIBILITY in 1 and ASSESSMENT and COMPATIBILITY in 2. They contain the collocates listed in Table 3.

2. VP → Verb + NP
  - Verb → ASSESS
  - NP → DET + N
  - DET → the
  - N → COMPATIBILITY

ASSESS → assess, examine, consider, review, prejudge, enquire, appraise,  
verify

COMPATIBILITY → compatibility

3. NP → N + PP

N → ASSESSMENT

PP → P+ DET + N

P → of

DET → the

N → COMPATIBILITY

ASSESSMENT → assessment, examination, analysis, review

COMPATIBILITY → compatibility

A further investigation shows that both structures further colligate with the prepositional phrase PP+DET+N. In [4] first the general phrasal structures and then the local grammar categories are described. The nouns that occur in this prepositional phrase can be classified according to their denotation into three classes: legal acts (coded as LEGAL ACT), international companies (coded as COMPANY) and financial support (coded as AID).

4. PP → P + DET+ N

P → of

DET → the, a, that, zero plural

NP → LEGAL ACT, COMPANY, AID

LEGAL ACT → legislation, right, rule, decision, decree

COMPANY → concentration, merger

AID → aid, measure, transaction

The data in the present study allow further specification of the semantic and grammatical profile of *compatibility* [5]. The existing structure colligates with the prepositional phrase PP+DET+N. The preposition observed here is *with*, the determiner is either *the* or empty plural and the types of nouns observed are conditioned by the items established at the previous stage. The items from the categories AID and COMPANY collocates only with *common market* (coded as COMMON MARKET), whereas the items from the category LEGAL ACT collocates with expressions that refer to EU law (coded as EU LAW).

5. PP → P + NP

P → with

NP → DET+N

DET → the, empty plural

N → COMMON MARKET| EU LEGAL DOCUMENT

COMMON MARKET → common market

EU LAW → Treaty, Second Directive, EU or Community Law

With this, the final stage of the analysis of the local grammar associated with *compatibility* is reached. [6] shows the complete structure in terms of parts of speech categories. LU here refers neutrally to the whole construction as a linguistic unit. The entire local grammar is displayed in the form of a finite-state graph in Figure 4. Subscripts indicate which types agree with each other. For example, AID<sub>1</sub> and COMPANY<sub>2</sub> agree with COMMON MARKET<sub>1</sub> but not with LEGAL ACT<sub>3</sub>. Vertical bars again indicate alternatives.

6.

$$\text{LU} \rightarrow \begin{matrix} \text{VP} + \\ \text{NP} + \end{matrix} \left[ \begin{matrix} \\ \text{PP} + \text{PP} \end{matrix} \right]$$

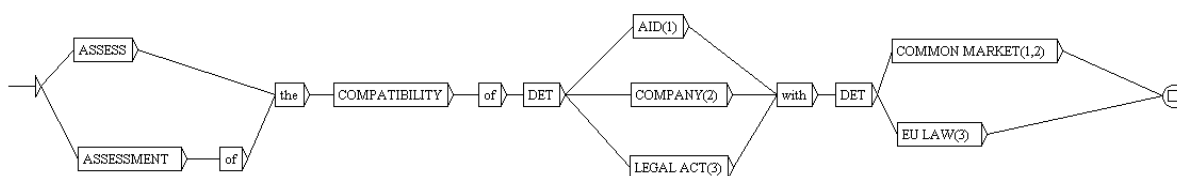


Figure 4: A local grammar diagram of *compatibility*

This analysis demonstrates that CJEU judgments contain hybrid expressions which result from the translation of judgments from French into English. As illustrated through the analysis of *compatibility* such hybrid expressions have a unique grammatical and semantic profile that cannot be found in the standard UK English varieties. A repeated use of the same lexical items with *compatibility* indicates also the formulaic nature of this hybrid expression.

The results of this analysis also demonstrate that a combination of a keyword analysis with the investigation of words that have origin in a specific language can help identify potential candidates

for hybrid expressions. Whether a candidate is a hybrid expression or not is then established through a comparative analysis collocation analysis. Finally, a local grammar analysis provides a detailed description of the types of structures, patterns and items associated with a given hybrid expression.

## **7. Conclusion**

The following conclusions about the nature of CJEU judgments follow from the above studies:

- The high degree of formulaicity is one of defining features of CJEU judgments;
- The argumentation of CJEU judgments relies on a limited number of textual devices;
- CJEU judgments contain hybrid expressions which are created through translation;

Furthermore, the studies prove the validity of the models proposed in the present paper. The first study demonstrates how the degree of formulaicity can be studied at the textual level. The second study illustrates that an investigation of sentence-initial textual Themes can show how these expressions signal the development of information in texts. Finally, the local grammar approach provides a fine-grained description of grammatical and semantic structures of hybrid expressions.

In a previous study (Trklja and McAuliffe, forthcoming) it was demonstrated that paragraph initial multi-word units signal the discourse organization of the entire texts of CJEU judgments. The main pattern observed in that study was that the argumentation is based on the Consideration-Conclusion pattern. Semantically, this pattern corresponds to the consequential type of causal-conditional types of logico-semantic relations. The lexical items that signal this type of relations also occur with high frequency in the position of textual Themes. It follows that the same kind of devices are used as discourse organizers both at the macro level of entire texts of judgments and at the paragraph level. These relations serve as the primary principle of argumentation in CJEU judgments. Following Koestler (1964) it can be argued that these devices indicate routinization of thinking at the CJEU. Routinization is understood as the process of selection “the sub-codes of grammar and syntax ... [which are] are almost wholly automatized” (Koestler, 1964: 12).

One might wonder how the evidences of routinized thinking may be reconciled with the findings that demonstrate the use of hybrid expressions. Hybrid expressions are associated with the creation of new cultural forms, undermining of the established ways of thinking and variety. This question can be answered only briefly here. First, although it is true that translation creates semantic

diversity in CJEU judgments due to re-selection of established translation candidates it also serves as a force that controls that this diversity does not devolve into chaos and disintegration. Second, once new concepts have been created they become established and through repetitive and routinized reasoning they play an important role in embedding rule of law. Through its case law, the early CJEU developed and extended its own jurisdiction and transformed the European Union from a traditional international organisation into a new type of legal order (Harmsen and McAuliffe, 2014). As the EU legal order became more established, the level of lexical variation in CJEU judgments seems to have dropped.

## Acknowledgments

The research for this paper was carried out as part of the European Research Council (ERC) funded project ‘Law and Language at the European Court of Justice’. For more details of this project please see [www.llecj.karenmcauliffe.com](http://www.llecj.karenmcauliffe.com). I wish to thank anonymous reviewers, the editorial team of the book and Karen McAuliffe for constructive and helpful comments on the earlier version of the paper. The usual disclaimers apply.

## References

- Bakhtin, M. M. (1981) *The dialogic imagination: Four essays* (C. Emerson, Trans., M. Holquist, Ed.). Austin, TX: University of Texas Press.
- Bhabha, H. K. (1994) *The location of culture*. New York: Routledge.
- Biber, D. (2006) *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins Publishing.
- Biber, D. (2009) ‘A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing.’ *International Journal of Corpus Linguistics*, 14(3), 275-311.
- Biber, D. and Conrad, S. (1999) ‘Lexical bundles in conversation and academic prose.’ *Language and Computers*, 26, 181-190.
- Biber, D., Conrad, S. and Cortes, V. (2004) ‘If you look at...: Lexical bundles in university teaching and textbooks.’ *Applied linguistics*, 25(3), 371-405.
- Biel, Ł. (2014) *Lost in the Eurofog: The Textual Fit of Translated Law*. Frankfurt am Main: Peter Lang.
- Bond, N. (2001) ‘Interpreting the objectively “strange” and the strangely “objective”.’ *Across Languages and Cultures* 2(2): 251–259.



Born, J. S. W. (1995) *Eurotexte: Textarbeit in einer Institution der EG*, Tübingen, Gunter Narr.

Brenner, S. (1992) *Precedent Inflation*. New Brunswick, NJ: Transaction Publishers

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

Chomsky, N. (1957) *Syntactic structures*. New York: Mouton.

Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge: MIT Press.

Conklin, K. & Schmitt, N. (2012) 'The processing of formulaic language.' *Annual Review of Applied Linguistics*, 32, 45–61.

Crombie, W. (1985) *Process and Relation in Discourse and Language Learning*. Oxford: Oxford University Press.

Dalton-Puffer, C. (1996) *The French Influence on Middle English Morphology: a Corpus-Based Study on Derivation*. Berlin: Walter de Gruyter.

Fries, P. H. (1981) 'On the status of Theme in English: Arguments from discourse.' *Forum Linguisticum*, 6(1), 1-38.

Fries P. H., (1995) 'A Personal View of Theme.' In Ghadessy M. (ed.), *Thematic Development in English Texts*, 1-19, London: Pinter. pp 1-20.

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

Gross, M. (1987) 'The use of finite automata in the lexical representation of natural language.' *Lecture Notes in Computer Science 377*, Springer-Verlag.

Gross, M. (1993) 'Lexicon Based Algorithms for the Automatic Analysis of Natural Language.' *Theorie und Praxis des Lexikons*. Berlin: Walter de Gruyter.

Gross, M. (1997) 'The Construction of Local Grammars.' In *Finite-State Language Processing*. Cambridge, Massachusetts: The MIT Press, pp. 329–354.

Halliday, M.A.K. (1985) *An Introduction to Functional Grammar*. London: Edward Arnold.

Halliday, M.A.K. (1991) 'Towards Probabilistic Interpretations.' In: Ventola, E. (ed.) *Functional and Systemic Linguistics: Approaches and uses*. Berlin and New York: Mouton de Gruyter, pp. 39–61.

Hannerz, U. (1987) 'The world in creolisation.' *Africa*, 57(04), 546-559.

Harris, Z.S. (1954) 'Distributional Structure.' *Word*, 10(2/3), 146-162.

Harris, Z. S. (1988) *Language and information*. New York: Columbia University Press.

- Harris, Z. S. (1991) *A Theory of Language and Information: A mathematical approach*. Oxford & New York: Clarendon Press.
- Harris, Z. (2002) 'The structure of science information.' *Journal of Biomedical Informatics* 35(4), 215-221.
- Harmsen, R and McAuliffe, K. (2014) 'The European Courts.' In Magone J.M. (ed.) *The Handbook of European Politics*. London: Routledge.
- Hobbs, J. R. (1985) 'On the coherence and structure of discourse.' *Technical Report* 85-37, Center for the Study of Language and Information (CSLI), Stanford, CA.
- Kermas, S. (2010) 'English Legal Discourse and the French Continuum.' In Giannoni D. and Frade, C. (eds.), *Researching Language and the Law*. Berna: Peter Lang, pp. 49-69.
- Koestler, A. (1964) *The act of creation*. London: Hutchinson.
- Martin, J. R. (1992) *English text: System and structure*. Amsterdam: Benjamins.
- Martin, J.R. (1995) 'More than what the message is about: English Theme.' In Ghadessy, M. (ed.) *Thematic Development in English Texts*. London: Pinter, pp 223-259
- McAuliffe, K. (2009) 'Translation at the Court of Justice of the European Communities.' In: Oslen, F and Stein, D. *Translation Issues in Language and Law*. New York: Palgrave Macmillan.
- McAuliffe, K. (2011) '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, 97-115.
- McAuliffe, K. (2013) 'The Limitations of a Multilingual Legal System'. *International Journal for the Semiotics of Law*, 26 (4), 861-882.
- Montolío, E. (2001). *Conectores de la Lengua Escrita*. Barcelona: Ariel Practicum.
- Moore, N. (2016) 'What's the point? The Role of Punctuation in Realising Information Structure.' *Written English Functional Linguistics* 3/6, 1-23 .
- Muhr, R. and Kettemann, B. (eds.) (2002) *Eurospeak: der Einfluss des Englischen auf europäische Sprachen zur Jahrtausendwende*. Frankfurt am Mein: Peter Lang.
- Nesi, H. and Basturkmen, H. (2006) 'Lexical bundles and discourse signalling in academic lectures.' *International Journal of Corpus Linguistics*, 11(3), 283-304.
- Neubert, A. (2001) 'Some implications of regarding translations as hybrid texts.' *Across Languages and Cultures* 2 (2), 181-193.
- Pawley, A. and Syder, F. H. (1983) 'Two puzzles for linguistic theory: Nativelike selection and nativelike fluency.' In Richards, J.C. and Schmidt, R. W. (eds.) *Language and Communication*. London: Longman, pp. 191-225.
- Roche, E., and Schabes, Y. (1997) *Finite-state language processing*. MIT press.

Schäffner, C. and Adab, B. (2001) 'The idea of the hybrid texts and translation: contact as conflict.' *Across Languages and Cultures*, Vol 2, pp. 167-180

Schmitt, N. (Ed.). (2004). *Formulaic sequences: Acquisition, processing, and use*. Amsterdam: John Benjamins Publishing.

Sinclair, J.M. (1991) *Corpus, Concordance, Collocation*. Oxford: OUP.

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

Tirkkonen-Condit S. (2001) 'EU Project Proposals as Hybrid Texts: Observations from a Finnish Research Project.' *Across Languages and Cultures*, 2 (2), 261-265.

Trklja A. and McAuliffe K, (forthcoming) 'Metadiscursive signalling devices in legal language: A corpus-based model for studying discourse organisation.' *Journal of Applied Linguistics*.

Wray, A. (2005) *Formulaic language and the lexicon*. Cambridge University Press.

Zergollem-Miletic, L. (1997) 'Morphological adaptation of the suffixes of English nouns borrowed in French.' *Studia Romanica et Anglica Zagrabienia* 42, 411–416.

Young, L. (2000) 'Hybridity's Discontents: Rereading Science and "Race".' In Brah, A. and Coombes, A. (eds.) *Hybridity and its Discontents: Politics, Science and Culture*. London: Routledge. pp. 154-170.