Edyta Więclawska

University of Rzeszów, Poland

e-mail: edytawieclawska@interia.pl

ORCID: 0000-0003-0798-1940

QUANTITATIVE DISTRIBUTION OF VERBAL STRUCTURES WITH REFERENCE TO THE AUTHORSHIP FACTOR IN LEGAL STYLISTICS

Abstract. The paper aims at describing the findings and conclusions formulated in the analysis of the authorship factor in legal discourse. It is hypothesised that verbal structures show systemically varied distribution across legal discourse and the relevant distinctions run through the authorship categories. When it comes to the aim of the research it draws on the tradition of sociolinguistic methodology targeting issues related to language variation which follows the basic assumptions of functional grammar. From the point of view of the material covered by the analysis it contributes to the research on legal discourse and specifically on its specialised domain referred to as corporate, company or business discourse. It provides additional empirical data pointing to the non-homogeneity of the legal style and formal distinctions originating from rich contextual background. The study is conducted on the material of a custom-designed corpus of English legal texts, classified as secondary genres.

Methodologically, the study makes use of the tenets of supervised search of digitalised corpora and automatic data extraction based on discrete units, subsequent identification of recurring longer contiguous and/or non-contiguous sequences, if any, built around the axis of specific verbal structures and finally qualitative comparative analysis (characterisation) of the material. The discussion presents sample data and focuses on the most salient categories, both quantitatively and qualitatively. The inductive approach confirms the formal divergencies in the communicative situation covered by the analysis.

The findings encapsulate patterns and tendencies in the quantitative distribution of verbal structures depending on the authorship category. It may be concluded that authorship is a factor delineating distinctions as regards (i) the repertoire of grammatical instruments exploited (verbal structures), which contributes to the specific stylistic profile of given authors. This shows that the thesis posed is verified positively and the study shows further, more detailed distinctions running through groups of subcategories distinguished within the authorship categories specified upon the start of the research.

Keywords: stylometry, verbal structures, authorship factor, stylistic distinctions, linguistic variation, legal discourse, institutional setting, professional title, supervised search, corpus analysis, comparative analysis.

1. Introduction

The analysis discussed in this paper fits in the strand of linguistic research which aims at identifying stylistically distinctive areas within various specialised registers, legal language being one of them. The findings that emerge from the relevant studies point to the non-homogeneity of legal discourse with distinctions running mainly through generic factors (e.g. Bhatia, 2014). Variations are further identified when relevant data are analysed against time- and culture-related axes (Peruzzo, 2017). The study indirectly draws on the assumptions found in corpus sociolinguistics (e.g. Romaine, 2009), and it relates to such concepts as identity issues (e.g. Dodsworth, 2014; Mesthrie, 2014), speech style variation, communicative divergence, group level identities and accommodation theory (e.g. Giles, 2009). With regard to the domain, it is hoped that the study makes a modest contribution to sociolinguistic studies concerning legal language (e.g. Bhatia, Bhatia, 2011; Innes, 2016), sociolinguistics and workplace discourse with a focus on corporate discourse (Bhatia, 2017, pp. 195–217; de Groot, 2014; Galdia, 2017, p. 93), discourse of business in large business organisations (e.g. Gunnarsson, 2009), business discourse in the globalised economy (Gunnarsson, 2009, pp. 220–235).

The issue of the authorship factor has been addressed in research studied from the highly technical perspective of authorship attribution (e.g. Bhargava, Mehndiratta, Asawa, 2013; Coyotl-Morales, Villaseñor-Pineda, Montes-y-Gómez, Rosso, 2006) and it concerns various types of register. This study contributes to relevant findings in the area of legal language (e.g. Nirkhi, Dharaskar, 2013) and it presents an induction-based linguistic perspective, where specific motifs defined as contiguous strings of words forms, lemmas and parts of speech tags are found typical for the style of individual authors. A study of French novels may constitute an example in point (Legallois, Charnois, Larjavaara, 2018b, p. 165). In general, the findings emerging from such context-related perspectives allows us to construe a fairly coherent model of legal communication which evidences that legal language is admittedly not a homogenous phenomenon and sometimes the stylistic, grammatical rules are vague and difficult to capture but it is far from chaotic, disorganised and unpredictable. The distinctions are shared by the category of authors, institutional settings, generic conventions, to name just a few factors.

When it comes to methodological tradition the study primarily fits in the discrete unit approach, addressing extraction and analysis of distribution patterns of closed class items (Legallois, Charnois, Larjavaara, 2018a, p. 3), but in the post-extraction phase the discussion makes a modest contribu-

tion to the syntagmatic, sequential approach in linguistic analysis in that we focus on how the discrete units extracted from the corpus function in longer strings of syntactic structures which are found salient and author-specific (Legallois, Charnois, Larjavaara, 2018a, p. 3; Legallois, Charnois, Larjavaara, 2018b, p. 165; Longerée, Mellet, 2018, p. 146). By virtue of their quantitative overrepresentation linguistic peculiarities of the style of specific authorship categories are found.

The research attempts to identify some further empirical evidence legitimising the basic assumptions of functional grammar. It addresses the issue of the acceptability and communicative effectiveness of a multitude of formal realisations, along the functional grammar paradigm acknowledging the coexistence of systemic and cultural forms. These concepts have been discussed by the father of functional grammar (Halliday, 2004) and his followers (Lewiński, 2013).

Specifically, the research addresses the issue of stylistic distinctions in the realm of legal discourse in that it aims at identifying patterns attributable to specific authorship categories on the material of verbal structures¹. It needs to be noted here that the concept of authorship is understood from the collective perspective as a category encompassing the stylistic, grammatical and rhetorical specificities of a type of individual drafters of legal documents, these being qualified by virtue of their institutional affiliation and professional capacity. The questions posed by the author are as follows: 1. Are there any authorship-related patterns in the frequency distribution of verbal structures? 2. Do the distinctions, if any, enable further sub-categorisation in the authorship categories? 3. Can we identify any salient contiguous or non-contiguous sequences with quantitatively salient verbal structures acting as their components that are stylistically and rhetorically important and distinctive for the specific authorship categories?

It is hypothesised that the repertoire of verbal structures exploited by various categories of authors varies and that distinctions in the frequency distribution of verbal structures covered by the analysis allows us to identify some patterns. Furthermore, the hypothesis is that the verbal structures act as components of contiguous sequences of words, largely formulaic, which are carriers of the stylistic potential for given authorship categories.

2. Methodology

The corpus methodology applied in the study may be described with some of the conventional parameters and terminology adopted in linguistic corpus analyses. The study was conducted on a custom-designed corpus of

English documents, comprising commercial law documents, gathered in an exhaustive search of court files at two court divisions of the same type.² The language corpus used for this study is a part of larger set of texts making up a parallel English/Polish corpus and here it is used as a monolingual, thematic corpus. The English language corpus alone has 1,124,204 tokens, 932,839 words, 60,539 sentences and the lexicon size is calculated at the level of 25,935. The corpus has been pre-processed both manually and automatically in compliance with the frameworks for tagging and coded for the relevant metadata³ (Aijmer, 2009; Lehmberg, Wörner, 2009; Schmidt, 2009). The texts verbalise certain acts in law and/or confirm legal facts which are found specific for the said communicative setting. 11 categories of texts were identified in the corpus and these include: confirmation of registration, company extracts, foundation acts, declarations of will, financial documentation, report, authentication, authorisation, verification, resolution and miscellaneous. The corpus was annotated for context-related metadata and the ones which are significant for this analysis are the country of publication and the year of origin of the said texts. Geographically, the corpus texts were assigned to the category of text of American, British, European, Asian and Canadian origin. Diachronically, the corpus covers text spanning over the 60 year and here three categories were identified: < 2000, 2001–2010, > 2010.

For the extraction of the relevant verbal formula it was processed automatically with the CQL formula by the lexical processing programme Sketch Engine. The data were subsequently computed statistically in the R tool to demonstrate quantitative distinctions.

It needs to be noted that the study satisfies the basic recommendations and vistas for corpus studies on specialist languages as currently voiced days and these advocate (i) analysing small, highly specialised specialist corpora, that is corpora related to a specific communicative situation instead of constituting a general account of a specialist register which tend to fail to capture intra-register distinctions, (ii) examining extralinguistic factors in legal communication, which is possible to achieve effectively when the texts making up the corpus represent one type of legal discourse but at the same time exploit the contextual and generic range of the communicative situation.

For the purpose of analysing the authorship factor in the most comprehensive way the author took into account two levels on which the authorship factor can be examined. The first dimension covers authorship categories identified at the level of the institution to which an author is affiliated (**INSTITUTIONAL NAME**). It is assumed that the in-

stitutional background greatly affects the drafting style (cite). These involve ENTITY ENTERED INTO THE REGISTER, AUTHENTICATION AUTHORITY, COMPANY REGISTRATION AUTHORITY, COMPANY EXTERNAL ENTITY PROVIDING PROFESSIONAL SERVICES NOT CLASSIFIED ELSEWHERE, MISCELLANEOUS. The second dimension of the analysis covers authorship categories discerned with the realm of **PROFESSIONAL TITLE** and these are as follows: ENTITY ESTABLISHING THE COMPANY, COMPANY MANAGER, COMPANY OFFICER, ENTITY PARTICIPATING IN THE COMPANY, ENTITY AUTHORISED TO REPRESENTATION, NOTARISATION OFFICER, FOREIGN SERVICE POST, STATE CERTIFICATION AND LEGALISATION AUTHORITY, HEADS OF REGISTRATION AUTHORITY, OFFICERS OF REGISTRATION AUTHORITY OF LOWER LEVEL, LEGAL COUNSEL, TAX AUTHORITY.⁴ The quantitative data related to nine verbal structures, extracted from the corpus were statistically processed and the frequency distribution patterns for the individual authorship categories within the domain of professional title and institutional name were compared. The choice of the verbal structures scrutinised include the following, starting from the top, reading Figures 1, 2 and 3 in a clockwise direction: *modals with past reference followed by active infinitive, modals with present reference followed by active infinitive, modals with present reference followed by passive infinitive, present perfect active forms, present perfect passive forms, simple past active forms, simple past passive forms, simple present active forms and simple present passive forms*. The verbal structures covered by the analysis exhausted the repertoire of the verbal structures used in the texts. No other forms were identified in the random, sample analysis conducted in the pre-processing stage.

The discussion aims at presenting sample data that are illustrative for the most salient and recurring quantitative distinctions, as found representative for the corpus in general. It proceeds from discussing the material for the authorship categories by analysing one comparative set of data at the institutional level, as found illustrative for the whole corpus in view of its quantitative and cognitive salience. Subsequently, the discussion continues to present the most striking contrasts along the authorship factor axis of professional titles. Since the analysis unveiled significant and consistent distinctions running through groups of authorship categories within the professional title framework, in this part of the discussion the author presents two groups of compared data sets. After commenting in brief upon the wholistic frequency-related contrasts of the juxtaposed data sets, se-

lected according to the criteria specified above, the discussion focuses on one verbal structure which is found to be quantitatively and qualitatively most significant. Statistically overrepresented features of a language are assumed to have the status of ‘linguistic singularities peculiar to an author’ (Legallois, Charnois, Larjavaara, 2018, p. 164). This is considered to legitimise closer discussion of the top frequencies. The verbal structures subject to analysis are discussed on the basis of concordance data and they are studied for some patterns with regard to their immediate context, variation potential and stylistic function capacity. The five top most hits are discussed in detail (threshold of 5).

3. Results

3.1 Authorship factor at the level of INSTITUTIONAL NAME

The tendencies in the distribution of verbal structures with regard to the authorship factor against the framework of institutional categories are illustrated on the material corresponding to two entities. These include the categories labelled as AUTHENTICATION AUTHORITY and ENTITY ENTERED INTO THE REGISTER. The criteria for selection of these authorship categories for discussion are as follows: (i) they illustrate recurring distinctive patterns with regard to the quantitative distribution and are thus expected to be discursively distinct, and the quantitative distinctions here are found to be discriminative also for other categories within the framework of institutional name, and (ii) they constitute interesting material for analysis in that they represent two distinct types of drafters of legal instruments if we apply the criterion of degree of professionalism. Comparing authorship categories that vary in the professional provenience should ensure interesting material for analysis of the rhetorical potential of the language use. To explain, we have drafters of legal documents who are assumed to be actors in legal communication without professional qualifications, representing the so-called lay scene in legal communication, the category labelled ENTITY ENTERED INTO THE REGISTER being an example here. The other category selected for comparative analysis is assigned the label AUTHENTICATION AUTHORITY and comprises agents using professional style in legal communication, observing the terminological regime at the syntagmatic and paradigmatic level.

Figure 1 presenting a comparative account of the distribution of verbal structures across two authorship categories shows a few salient areas of overrepresentation.

**Figure 1. Authorship factor at the level of INSTITUTIONAL NAME:
Comparative account of distribution of verbal structures**

Present perfect active forms are selected for closer discussion here on the grounds of being quantitatively salient in some of the authorship categories and by virtue of being important tools to materialise important communicative functions of the legal documents.

Present perfect tense active voice shows as overrepresented in the authorship category AUTHENTICATION AUTHORITY with a result of 138/122.75 per million. The loser in the dataset contrasted in Figure 1, that is ENTITY ENTERED INTO THE REGISTER, is registered with a score of 1,034/919.75 per million. The five top hits for present perfect active forms in the winning authorship category include the following entries: *have subscribed* 18 (16.01 per million), *have signed* 18 (16.01 per million), *have hereunto set* 17 (15.12 per million), *have caused* 10 (8.90 per million), *have put* 7 (6.23 per million).

With regard to the lexical verbs used in the present perfect active structures they all denote various formal measures used to validate a document or effectuate acts in law confirmed by a document and protect it against any fraudulent transmission in the legal trade. With the exception of the causative *have caused* (EXTRACT 4) which also corresponds to a specific rhetorical convention of legal discourse, all the lexical verbs denote individual activities conventionally preformed with the aim to legalise a document – putting a signature on a document (EXTRACTS 1, 2, 3 and 5). This brings us to the issue of the performative and iconic function of legal language which presupposes the use of specific formula for effective enforcement of acts in law (Witczak-Plisiecka, 2007).

EXTRACT 1: #105490 • UK • 192 • 2 • 014 • 30 • 6 • K • 2 • 2013 • 7
• 2 • 2013 • 1 • 131⁵

*IN FAITH AND TESTIMONY whereof I the said Notary **have subscribed** my name and set and affixed my seal of office at Birmingham aforesaid this 13th day of November 2013.*

EXTRACT 2: #29150 • UK • 42 • 2 • 004 • 1 • 6 • K • 1 • 2004 • 7 • 1 • 2004 • 1 • 74

*IN FAITH AND TESTIMONY whereof I the said Notary **have signed** my name and set and affixed my Seal of office at Uxbridge England.*

EXTRACT 3: #169000 • UK • 283 • 2 • 022 • 6 • 6 • K • 2 • 2006 • 7 • 2 • 0 • 3 • 210

*IN TESTIMONY WHEREOF I **have hereunto set** my hand and affixed my Seal of Office in the City of London, aforesaid, this fifteenth of May.*

EXTRACT 4: #159164 • UK • 264 • 2 • 022 • 10 • 6 • K • 2 • 2007 • 7 • 2 • 0 • 3 • 205

*THAT I **have caused** an inspection to be made at the Companies Registration Office for England and Wales of file No.*

EXTRACT 5: #682865 • UK • 903 • 2 • 071 • 8 • 6 • M • 2 • 2014 • 7 • 2 • 2014 • 1 • 148

*IN FAITH AND TESTIMONY hereof I the said notary **have put** my name and affixed my seal of office at London this 14 March 2014.*

The illustrative material related to the winning authorship category allows us to formulate some further findings regarding the formulaicity potential of the verbal structures analysed. The very verbal formula extracted from the corpus prove to operate as components of longer strings of repetitive structures. This confirms the highly formulaic character of legal language as regards the analysed authorship category. The present perfect active structures are embedded into a ritual expression which starts with (*in faith and testimony*) and it is occasionally followed by the emblematic pro-forms (*hereunto, hereof*). Secondly, the material presented in EXTRACTS 1–3 and 5 shows that formulaic sequences including a verb in present perfect active form are fairly resistant to variation against contextual text variantivity. The extracts quoted above illustrate only sample material and they prove to be recurring as prefabricated chunks, both in the diachronic and diatopic perspectives throughout the corpus, as is supported by the varied time allocation of the extracts.

At the same time it needs to be noted that the data exemplified in EXTRACTS 1, 2, 3 and 5 testify to there being a variantivity potential as

regards linguistic expression of the document legalisation function. The material quoted above and some additional instances extracted from below of the threshold top five (e.g. *have subscribed*, *have personally signed*) evidence the admissibility of many alternative ways successfully operating in one communicative setting, which is consistent with the communicative flexibility approach advocated by functional grammar. The patterns that may be identified in the range of conventions include, for example, employment of lower register forms (e.g. *put signature* or *make signature* instead of *sign*) or adding the ritual pro-forms (e.g. EXTRACT 3 *hereunto*).

3.2. Authorship factor at the level of PROFESSIONAL TITLE

The tendencies observed while analysing the authorship categories fitting in the dimension of **PROFESSIONAL TITLE** justify organising the discussion around two sets of data. The first set of data covers selected authorship categories classified as lay agents in legal communication (Figure 2), while the second data sample features the professional style in legal communication (Figure 3). The feature of professionalism is ascribed to the author on the grounds of having professional qualifications in the area of law.

Simple past active voice forms constitute one of the areas where a quantitative hiatus is noted between the two non-professional authorship categories contrasted. Simple past active voice structures are also selected for closer discussion on the grounds of their stylistic significance. The category of COMPANY MANAGER and COMPANY OFFICER are found representative for the said tendency. Figure 2 presents visual representation of the relevant material.

Figure 2. Authorship factor at the level of PROFESSIONAL TITLE/lay users: Comparative account of distribution of verbal structures

Out of a total number of verbal structures in simple past active voice amounting to 5,053/4,494.74 per million, the texts authored by the entities

categorised as COMPANY OFFICER scored 1,662/1,1478.38 per million as compared to those authored by the entities categorised as COMPANY MANAGER scoring 78/69.38 per million. From the point of view of semantical profile the top of the list is dominated by the hits conceptually related to the domain of verbs standing for expression of will (e.g. *stated, declared, reported, adopted, noted, approved*), which shows individual stylistic preferences attributable to the said authorship category. The top 5 positions for the winning category (COMPANY MANAGER) cover the domain of temporal relations (*ended* scoring 46/40.92 per million), reference to legal provisions (*provided that* 39/34.69 per million and *referred to* 37/32.91 per million), and confirmation of participation (*was/were present* 35/31.13 per million) and acknowledging the capacity to act (*authorised* 36/32.02 per million).

As in the case of the material discussed above within the framework of INSTITUTIONAL NAME the data extracted here disguise the high level of recurrent sequences including the simple past active verb forms. A case in point is the verb *were present [and agreed to undertake]* which remains consistent across the distinct legal cultures, as evidenced by EXTRACTS 6 and 7 representing data from Finland and Canada.

EXTRACT 6: #558833 • Fin • 804 • 1 • 060 • 3 • 2 • 0 • 1 • 2012 • 10 • 1 • 0 • 0 • 94

*All Board members **were present** and agreed to undertake this resolution in written form.*

EXTRACT 7: #766851 • Can • 981 • 1 • 079 • 6 • 2 • 0 • 1 • 2015 • 10 • 1 • 0 • 0 • 91

*All Board members **were present** and agreed to undertake this resolution in written form.*

The high level of repetitiveness of specific contiguous, verb embedding strings confirms the strongly conventional and schematic character of legal language within specific authorship categories, which is found to recur through different texts and different contextual backgrounds. The recurring, prefabricated sequences include [*directors*] *were present and acting throughout [...], [...] was present and the meeting was ready to proceed [...], [...] a quorum was present according [...], [...] were present and agreed to undertake a resolution [...]*. Other examples include sequences with the simple past verbal component *authorised* include [...] *authorised to represent/individual representation [...], [...] authorised to conduct [...], [...] authorised to act and sign on behalf of [...], [...] authorised to call special meeting [...]*.

Notably, the quantitatively salient verbal material featuring the authorship category COMPANY MANAGER proves to be featured by coordinate structures of the binomial type (Groover, 1999). Binomials have always been a feature of legal language (Kopaczyk, 2013, pp. 188–206) and they make up a strong characteristics of legal language contemporarily too (Giammarresi, 2010; Jopek-Bosiacka, 2006, p. 63). They are considered to be a kind of rhetorical figure typical for ritualised formal language (Sauer, 2017). In our case the examples extracted from the corpus, including those ranked below the threshold of 5 top hits include *reacquire and retire*, *prepare and post*, *pay or set*, *pay, declare or announce*, *pay, declare, issue or sell*, *issue or deem*, *issue and send*, *form and register*, *declare and announce*, *declare and pay*, *ask and take*. The area of legal communication delineated by the authorship category noted the instances of the same binomials used across various legal cultures. EXTRACT 8 exemplified material from the UK, EXTRACT 9 – Cyprus and EXTRACT 10 – US. The binomials extracted from the corpus exemplify various classifications capturing various lexical categories. The categorisation of the relevant corpus material involved identification of the function-oriented category covering expression of appraisal (*issued or deemed issued*), instances of various categories relating to the semantic motivation of the coordinate units (e.g. binomials conjoined by the antonym – *lent or advanced*).

EXTRACT 8: #1011186 • UK • 1246 • 1 • 106 • 5 • 0 • K • 5 • 2016 • 11 • 5 • 2017 • 1 • 6830 * — (1)

*Subject to the articles, anything **sent or supplied** by or to the company under the articles may be sent or supplied in any way.*

EXTRACT 9: #468189 • Cypr • 656 • 4 • 052 • 11 • 11 • M • 6 • 2010 • 3 • 6 • 2010 • 1 • 12700

*exchange, and other negotiable instruments, and all receipts for monies paid to the Company, shall be **it accepted, endorsed**, or otherwise executed, as the case may be, in such manner as the Directors shall from time to time by resolution.*

EXTRACT 10: #966117 • US Cal • 1194 • 1 • 101 • 11 • 2 • K • 5 • 2010 • 3 • 5 • 2016 • 1 • 9157

*Shares of Common Stock **issued or deemed issued** as a result of a decrease in the effective conversion prices of any Preferred Shares.*

The authorship categories encompassed in the framework of **PROFES-SIONAL TITLE** deserve to be discussed with another contrastive set of

data in view of the distinct background of the two authorship categories contrasted and – at the same time – distinctions observed which are different from those observed in the case of non-professional authors. Data included in Figure 3 provide a consistent representation of the corpus by pointing to a specific tendency, testifying on the one hand to significant symmetry in the frequency distribution of the said verbal structures and – on the other hand – to some divergencies. The symmetry can be accounted for by the professional status they share. This shows that the scale of the stylistic distinctions running through the authorship categories varies and there are groups of categories that display less significant divergencies in the distribution of the verbal categories covered by the analysis. Professional drafters of legal texts consciously operate legal terminology, consistently resort to using prefabricated expressions, have awareness of the necessity to stick to the principle of intertextuality, which results in producing highly repetitive and symmetrical patterns explicit in meaning, clear to follow and generating the performative effect, as prescribed.

Figure 3. Authorship factor at the level of PROFESSIONAL TITLE/non-lay users: Comparative account of distribution of verbal structures

Modal present active forms were selected for more detailed analysis here on the grounds of their cognitive and quantitative salience in legal texts in general (Williams, 2005, pp. 114, 121) and by virtue of them being a quantitatively significant category of verbal structures that distinguishes the two authorship categories within the class of professional authors. The total score for modal present active forms in the corpus is 6,263/5,571.05 per million. As evidenced in Figure 3 the authorship category categorised as NOTARISATION OFFICER scored 104/92.51 per million compared to the loser categorised as HEAD OF REGISTRATION AUTHORITY, recorded with the score at the level of 10/8.9 per million. Closer analysis of the top

hits above the threshold adopted in this analysis shows that modals used in present active are dominated by *shall* and *may*, which is consistent with the findings gathered for legal language in general, irrespective of whether these involve prescriptive of secondary genres⁶. The only other modals used in present active involve quantitatively marginal instances of *would* and *should*, the first used consistently in adverbial clauses after the compound subordinating conjunction (Quirk, Greenbaum, 1973, p. 113) *except where* (EXTRACT 12) introducing a hypothetical situation. *Should* is used sparingly in the corpus and, as noted above, it constitutes a small percentage of all finite verbal constructions in prescriptive texts (Williams, 2005, p. 128) and this pattern shows to hold true for secondary genres, as confirmed by the corpus statistics. In the documents drafted by the agents categorised as NOTARISATION OFFICER *shall* was recorded in just one context in its deontic capacity to indicate a specific course of action, provided that the speaker considers it right in a given situation (EXTRACT 13).

EXTRACT 11: #220465 • UK • 351 • 2 • 025 • 7 • 6 • K • 6 • 2004 • 7 • 6 • 2007 • 1 • 10458

<s>as a result of the Qualifying Issue in question or any subsequent Qualifying Issues under this article except where it **would defeat** the purpose of this sub-article a reference in these articles to Ordinary Shares shall <s>

EXTRACT 12: #265366 • UK • 393 • 2 • 031 • 11 • 6 • K • 1 • 2007 • 7 • 1 • 2008 • 1 • 40

*Certification to be signed in the presence of a notary public who **should attest** to the identity of the signatories and their authority to represent Partners.*

The sequences (modal verbs together with the lexical verb that follows) registered above the threshold of 5 top hits include: *shall come* 14/12.45 per million, *may come* 7/6.23 per million, *shall take* 4/3.56 per million, *shall apply* 4/3.56 per million, *may require* 4/3.56 per million. The dataset composed of the five top-most hits includes instances of a modal being a component of a longer ritual formulaic sequence used by a notary public legitimising the authenticity of the document (EXTRACTS 13, 14, 15). Importantly, the formulaic sequences prove to be recurring in the diachronic perspective (EXTRACTS 13 and 14) and thus proves to be change resistant in the short diachronic perspective. For confirmation compare the third but last metadata in the said extracts, that is ‘2004’ and ‘2015’ respectively. They evidence that the same discursive properties and contextual profile

of *shall* are noted in the two time slots for which the corpus texts were coded (2001–2010 and > 2010).⁷

EXTRACT 13: #60889 • UK • 107 • 2 • 004 • 8 • 6 • K • 2 • 2004 • 7 • 1 • 2004 • 1 • 105

*TO ALL to whom these presents **shall** come.*

EXTRACT 14: #849354 • UK • 1100 • 2 • 092 • 11 • 6 • M • 2 • 2015 • 7 • 1 • 2015 • 1 • 101

*TO ALL WHOM these presents **shall** come, I, GRAHAM PHILIP PENN of Crawley, West Sussex, NOTARY PUBLIC duly admitted hereby state the following.*

EXTRACT 15: #678337 • UK • 889 • 2 • 071 • 19 • 6 • K • 2 • 2014 • 7 • 2 • 2014 • 1 • 141

*TO ALL TO WHOM these presents **may** come, I, EDWARD YOUNG, Notary Public, of 10 Carlos Place London W1K 3 AT, England duly certify the authenticity of this signature.*

The semantics of the winning classes of modal verbs, that is *shall* and *may*, shows as fitting in another consistent pattern too. Judging by the immediate linguistic context and lexical verbs with which the said modals cooccur, the primary meaning that they convey is to confer discretionary power of the agents, granted under statute, contract or via delegation, examples being found above the threshold of top 5 hits and below. Cases in point involve the sequences *shall apply*, *shall require*, *may direct*, *shall use*. The material quoted again clearly confirms the high stability and recurrence of longer sequences including modals. Sequences [*as occasion shall*] or *may require* show as recurring in two distinct texts authored by entities categorised as NOTARISATION OFFICER, even when these operate within distinct legal cultures, Canada and UK (EXTRACTS 16 and 17 below).

EXTRACT 16: #660906 • UK • 878 • 2 • 070 • 8 • 6 • K • 2 • 2014 • 7 • 2 • 2014 • 1 • 163

*Notary have granted these Presents under my Notarial Firm and Seal of Office, to serve and avail as occasion shall or **may** require.*

EXTRACT 17: #765207 • Can • 976 • 2 • 079 • 1 • 6 • K • 2 • 2015 • 1 • 2 • 2015 • 1 • 179

*being requested I have granted the same under my hand and notarial seal of office to serve and avail as occasion shall or **may** require.*

Acknowledging the high level of prefabrication, repetitiveness and symmetry among the texts making up the corpus with regard to the choice and larger linguistic context in which the modal verbs are used, the data extracted allow us to identify some minor cases of variation. A good example in point is the case of the two top-most sequences *may come* and *shall come* (EXTRACTS 14 and 15), which prove to vary across legal cultures.

4. Conclusions

The discussion conducted on the basis of the corpus material allows us to present some findings, and these involve the following: (i) selected types of verbal categories covered by the analysis show as distinctive features of some authorship categories, (ii) the quantitative distinctions identified in the research make it possible to group the authorship categories discerned within the framework of **PROFESSIONAL TITLE** according to the scale and type of quantitative distinctions into subcategories belonging to professional vs. lay-communication and (iii) there are systemic tendencies in the use of the verbal structures related to the quantitatively salient areas. Specifically, the last point implies consistent use of prefabricated sentences with specific verbal structures as components, which is specific for a given authorship category and thus shows as its characteristic stylistic trait; formulaic capacity of the dominating verbal structures; and controlled variability in the structure of longer formulaic sequences with the specific verbal component.

The findings encourage us to draw conclusions with regard to the objective set to the analysis. The hypotheses related to the distinctiveness of the stylistic and grammatical profile of the authorship factor are positively confirmed. The scale in which the authors of the legal texts exploit verbal structures varies and the specific verbal structures are consistently used by the authors of legal texts in their formulaic capacity for the attainment of various communicative goals. This is found to profile the style of the specific authorship categories. In a somewhat larger perspective the study sets the ground and provides further arguments for us to claim that legal communication is far from chaotic and – in its diversity of structures and grammatical tools – the distinctions can be accounted for, among others, by the varied authorship of the texts. Verbal structures and the longer sequences they are components of prove to be a factor distinguishing various areas of legal communication and distinctions run through the linguistic performance of various authorship categories. In some cases – and the matter will have to be

analysed closer – the tendencies in the frequency distribution of the said discrete and non-discrete units is found to show further variations, depending on the contextual background and these distinctions can be accounted for functionally, on the grounds of the principle of intertextuality, the need to include prefabricated, ritual chunks to operate effectively for the attainment of a specific communicative purpose.

The findings gathered and conclusions formulated in this study encourage the formulation of some vistas as to how the research could be advanced further. Primarily, the findings could be verified by conducting more advanced cross-tabular analyses on the data available to exclude other contextual factors affecting the distribution of the data set achieved in the foregoing. Moreover, other types of statistical analyses could be carried out to verify the predictive capacity of the findings. Finally, as a natural course of action, a study should follow which would test the framework and thesis on different material with regard to the set of linguistic descriptors within the realm of discrete units and also outside of this category including contiguous and non-contiguous sequences and on different areas of legal discourse.

N O T E S

¹ For related analysis on the same set of verbal structures see (Więclawska, 2019b).

² The texts making up the corpus were extracted from the court files of 2 out of 21 divisions of National Court Register, Register of Entrepreneurs in Poland (Kraków and Rzeszów). The search is claimed to be exhaustive on the grounds of applying systemic search criterion and digitalizing all the texts on file for the year 2017 which meet the criterion of a case with a foreign element. For related analyses based on CorpCourt see, for example, (Więclawska 2019a, 2019b, 2020).

³ For more extensive description of the context categories coded for the texts making up the corpus see (Więclawska, 2020).

⁴ The concepts of professional title and institutional name cover more authorship categories compared to previous studies which were based on samples of the corpus used for this analysis (Więclawska, 2019a).

⁵ Each extract from the corpus is acknowledged with the metadata as assigned in Sketch Engine. These stand for the following in the order they are mentioned: token number, country in which the document is published, doc. id, institutional name, krs [court file number], krs_item [number of entry in the National Court Register], professional title, sex, source text wordcount, source text year, title, target text word count, target text year, type of translation, word count.

⁶ *May* and *shall* are found to be the most distinct markers of legal texts and are the only modals whose usage in legal texts is greater than in general usage (Foley, 2001, p. 193, after: Williams, 2005, p. 121).

⁷ For more in this see the Introduction.

REFERENCES

- Aijmer, K. 2009. Parallel and comparable corpora in A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics – An International Handbook*, 275–291. Berlin/New York: Walter de Gruyter.
- Bhargava, M., Mehndiratta, P. and Asawa, K. 2013. Stylometric Analysis for Authorship Attribution on Twitter in V. Bhatnagar and S. Srinivasa (eds.), *Big Data Analytics. Second International Conference, BDA 2013 Mysore, India, December 2013 Proceedings*, 37–48. New York/Dordrecht/London: Springer International Publishing.
- Bhatia, V.K. 2014. Analysing discourse variation in professional contexts' Communication in V. Bhatia and S. Bremner (eds.), *The Routledge handbook of language and professional communication*, 3–13. New York: Routledge Taylor and Francis Group.
- Bhatia, V.K. 2017. *Critical Genre Analysis. Investigating interdiscursive performance in professional settings*. New York: Routledge.
- Bhatia, V.K., Bhatia, A. 2011. Legal discourse across cultures and socio-pragmatic contexts. *World Englishes*, 30(4), 481–495.
- Coupland, N. (2014). Social Context, Style, and Identity in Sociolinguistics in J. Holmes and K. Hazen (eds.), *Research Methods in Sociolinguistics. A Practical Guide*, 290–304. Oxford: Wiley Blackwell.
- Coyotl-Morales, R.M., Villaseñor-Pineda, L., Montes-y-Gómez, M. and Rosso, P. 2006. Authorship Attribution Using Words Sequences in J.F. Martínez-Trinidad, J.A. Carrasco-Ochoa and J. Kittler (eds.), *Progress in Pattern Recognition, Image Analysis and Applications*, 844–853. New York/Dordrecht/London: Springer International Publishing.
- de Groot, E. 2014. Corporate communication and the role of annual reporting in S. Bremner and V. Bhatia (eds.), *The Routledge handbook of language and professional communication*, 220–236. New York: Routledge Taylor and Francis.
- Dodsworth, R. 2014. Speech communities, social networks, and communities of practice in J. Holmes and K. Hazen (eds.), *Research Methods in Sociolinguistics. A Practical Guide*, 262–275. Oxford: Wiley Blackwell.
- Foley, R. 2001. Going out in Style? Shall in EU legal English in *Proceedings of the Corpus Linguistics Conference 2001 University of Lancaster*, 185–195.
- Galdia, M. 2017. *Lectures on legal linguistics*. Frankfurt am Main: Peter Lang.
- Giammarresi, S. 2010. Formulaicity and Translation: A Cross-corpora Analysis of English Formulaic Binomials and their Italian Translations in D. Wood (ed.), *Perspectives on Formulaic Language*, 257–272. New York: Continuum International Publishing Group.
- Giles, H. 2009. The Process of Communication Accommodation in N. Coupland and A. Jaworski (eds.), *The New Sociolinguistics Reader*, 276–287. New York: Palgrave Macmillan.

- Groover, C. 1999. Coordination in K. Brown and J. Miller (eds.), *Concise Encyclopedia of Grammatical Categories*, 115–123. Amsterdam/Lausanne/New York/Oxford/Shannon/Singapore/Tokyo: Elsevier.
- Gunnarsson, B.-L. 2009. *Professional Discourse*. London/New York: Continuum International Publishing Press.
- Halliday, M.A.K. 2004. *An Introduction to Functional Grammar*. New York: M.A.K. Halliday and Christian Matthiessen.
- Innes, B. 2016. Summing up in jury trials as interactive discourse – one plank in the New Zealand judiciary effort to improve communication with juries in R. Lawson and D. Sayers (eds.), *Sociolinguistic Research Application and Impact*, 132–151. New York: Routledge Taylor and Francis Group.
- Jopek-Bosiacka, A. 2006. *Przekład prawny i sądowy*. Warszawa: Wydawnictwo Naukowe PWN.
- Kopaczyk, J. 2013. *The Legal Language of Scottish Burghs*. Oxford: Oxford University Press.
- Legallois, D., Charnois, T. and Larjavaara, M. 2018a. Grammar of genres and styles: an overview in D. Legallois, T. Charnois and M. Larjavaara (eds.), *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, 1–14. Berlin: De Gruyter Mouton.
- Legallois, D., Charnois, T. and Larjavaara, M. 2018b. The balance between quantitative and qualitative literary stylistics: how the method of “motifs” can help in D. Legallois, T. Charnois and M. Larjavaara (eds.), *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, 165–193. Berlin: De Gruyter Mouton.
- Lehmberg, T. and Wörner, K. 2009. Annotation standards in A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics – An International Handbook*, 484–501. Berlin/New York: Walter de Gruyter.
- Lewiński, P. 2013. Założenia gramatyki funkcjonalnej in A. Burzyńska-Kamieniecka and A. Libura (eds.), *Sapientia Ars Vivendi*, 459–475. Wrocław: Oficyna Wydawnicza ATUT.
- Longerée, D. and Mellet, S. 2018. Towards a topological grammar if genres and styles: a way to combine paradigmatic quantitative analysis with a syntagmatic approach in D. Legallois, T. Charnois and M. Larjavaara (eds.), *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, 140–163. Berlin: De Gruyter Mouton.
- Mesthrie, R. 2014. Analysing Sociolinguistic Variation in Multilingual Contexts in J. Holmes and K. Hazen (eds.), *Research Methods in Sociolinguistics. A Practical Guide*, 275–290. Oxford: Wiley Blackwell.
- Nirkhi, S. and Dharaskar, R. 2013. Comparative Study of Authorship Identification Techniques for Cyber Forensic Analysis. *International Journal of Advanced Computer Science and Applications*, 4(5), 32–35.
- Quirk, R. and Greenbaum, S. 1973. *A University Grammar of English*. Harlow: Longman Group UK Limited.

- Peruzzo, K. 2017. Legal system: an additional variable in the analysis of short-term diachronic evolution of legal terminology. *International Journal of Legal Discourse*, 2(2), 291–313.
- Ramnil, H., Panchoo, S. and Pudaruth, S. 2016. Authorship Attribution Using Stylometry and Machine Learning Techniques in S. Berretti, S.M. Thampi and P.R. Srivastava (eds.), *Intelligent Systems Technologies and Applications*, 113–127). New York/Dordrecht/London: Springer International Publishing.
- Romaine, S. 2009. Corpus linguistics and sociolinguistics in A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics – An International Handbook*, 96–112). Berlin/New York: Walter de Gruyter.
- Sauer, H. 2017. The World in Two Words: Binomials in two English Translations of the Lotus Sutra. *Linguistica Silesiana*, 38, 7–37.
- Schmidt, H. 2009. Tokenizing and part-of-speech tagging in A. Lüdeling and M. Kytö (eds.), *Corpus Linguistics – An International Handbook*, 527–552. Berlin/New York: Walter de Gruyter.
- Więclawska E. 2019a. Field-specific conventions in the translation of company law documentation for court proceedings. *Studies in Logic, Grammar and Rhetoric*, 58(71), 221–243.
- Więclawska E. 2019b. Sociolinguistic and Grammatical Aspects of English Company Registration Discourse. *Humanities and Social Sciences*, XXIV, 26(4).
- Więclawska E. 2020. Contextualising the Notion of Context in Jurilinguistic Studies. *International Journal for the Semiotics of Law/Revue Internationale de Sémiotique Juridique*. DOI: 10.1007/s11196-020-09701-0.
- Williams, C. 2005. *Tradition and Change in Legal English. Vernal Constructions in Prescriptive Texts*. Bern: Peter Lang.
- Witczak-Plisiecka, I. 2007. *Language, Law and Speech Acts Pragmatic Meaning in English Legal Texts*. Łódź: Wyższa Szkoła Studiów Międzynarodowych w Łodzi.