

Edyta WIĘCŁAWSKA¹

SOCIOLINGUISTIC AND GRAMMATICAL ASPECTS OF ENGLISH COMPANY REGISTRATION DISCOURSE

The paper examines the complexity of legal communication in English in the context of its stylistic variantivity. The research examines whether the variable of genre and selected sociolinguistic factors affect the grammar of legal texts, and, specifically the quantitative distribution of verbal structures. The analysis fits in the paradigmatic approach to generic and sociolinguistic studies, where the distribution of discrete units is presented in quantitative terms. The operationalisation of the research involved the identification of the relevant linguistic material in the corpus, its statistical processing, and conducting relevant R analysis to determine potential correlations. The findings showed that the legal texts making up the corpus were not stylistically homogenous; the varied provenance of the source texts (from a diatopic and diachronic perspective, and in terms of genre profile) affected the stylistic structure/grammar of the texts; the frequency distribution of same grammatical categories tended to be discriminative for various categories of texts. The research makes a contribution through addressing legal communication within the realm of secondary genres that are underrepresented in legilinguistic studies, constituting a significant yet problematic domain from an intra- and inter-linguistic perspective. The findings obtained in the analysis and patterns are of practical use in multi-national, institutional environments, where English is used as a lingua franca in corporate communication. Moreover, the conclusions drawn provide a foundation for studies on legal translation in professional settings, including the didactic perspective.

Keyword: legal language, sociolinguistics, stylistic variantivity, lingua franca, verbal structures.

1. INTRODUCTION

An overview of the research into legal language shows that over the last few years there has been an increasing tendency to foreground the phenomenon of variantivity² in text stylistics³. It is either that the researchers explicitly aim at identifying distinctions between various areas of legal communications, or that they conduct their analyses on various sections of legal language data, which is determined by the existing types of corpora, technical feasibility to process them and the market-related needs. Ultimately, the findings gathered

¹ Edyta Więclawska PhD, Institute of Neophilology, University of Rzeszów in Rzeszów; e-mail: edytawieclawska@poczta.fm. ORCID: 0000-0003-0798-1940.

² The term *variantivity* is used in the foregoing with reference to the potential of legal language to express one function with distinct stylistic/grammatical tools.

³ The empirical data caused that variation is recognised to be a feature typical also for specialised languages which is manifested in the development of new research areas such as Socioterminology and Sociocognitive Terminology. For more information (See: Peruzzo, 2017).

are either an account of differences between the dissected types of legal discourse or an account of the distinctive features of legal stylistics, held as universal for legal communication, yet annotated with the reservation that they were identified in a field-specific corpus. Whichever is the case it remains evident that legilinguistic research takes account of the intra-domain distinctions, and 'domain' here means the whole area of legal communication.

The factors causing distinctions in legal stylistics are manifold. The titles of the papers and monographs are often indicative of their nature. Hence, the stylistic distinctions run through (i) specific types of legal texts, (ii) specific institutional domains delineated by the types of users (in professional settings) or (iii) legal systems (in the EU context), (iv) specific national context (in Italian legal texts)⁴.

In the research under discussion the primary objective was to identify potential distinctions in the stylistic structure of the legal texts belonging to a communicative context that is specific and homogenous, yet varying in sociolinguistic and genre-related background. The study is intended to be a modest contribution to the corpus-based studies (See: for example, Baker, 2004; Pontrandolfo, 2015; Biel, 2014), making use of the concept of *shallow features*, also referred to as *discrete units* (See: Legallois, Charnois & Meri Larjavaara, 2018), processed in a supervised way (See: Longérée & Mellet, 2018). The author is aware of the shortcomings of the analyses based on the processing of discrete units, as voiced in the literature of the subject (See: Longérée & Mellet, 2018), yet the aim of this paper is to go somewhat beyond the basic model of this kind of analysis in which texts are thought to be mere material to be statistically processed and – firstly – take account of the sociolinguistic background of the texts and – secondly – possibly develop the qualitative description component. The variantivity factor will be measured by taking account of the variables of text genre, country origin of the documents and year of publication of the document. The choice of these variables may be said to be natural since the variables that were transformed in the categories of statistical description were found to be the factors differentiating the documents making up the corpus.

The added value of the study is also deemed to lie with the authenticity of the material, its representativeness of English as the lingua franca (See: House, 2013), its qualification as secondary genres⁵, the homogeneity of the material in terms of the legal domain the texts belong to (registration-related discourse) and application of apt and effective – as is deemed to be – querying techniques making use of detailed filtering against metadata selection and thus enabling presentation of detailed and precise results.

It remains the main objective of the study to confirm a well-established hypothesis about stylistic distinctions in the legal discourse. It is hypothesized that the texts making up the corpus which are related to the same communicative situation (company registration proceedings) do not make a stylistically homogenous group and the distinctions run through the variables causing the variantivity (i) genre, (ii) country of origin, (iii) year of the source texts.

⁴ Examples here include, in the order of mentioning: studies of court judgements (See: Pontrandolfo, 2015; studies on corporate discourse see: Bhatia, 2017; studies conducted on EU texts in general see: Biel, 2014; interlingual analyses – Ebeling, 1998.

⁵ Studies on secondary genres are claimed to be underrepresented in legilinguistic studies (See: Bhatia, 2006).

2. CORPUS AND DATA SELECTION; TASK OPERATIONALISATION

The analysis was conducted on the basis of a custom-designed corpus composed of English documents and their Polish translations, as found on court files at the National Court Register (working title: CorpCourt). The documents include commercial law documentation for court registration proceedings in Poland. The CorpCourt corpus as a whole covers source texts and their translations and has been used as a translation corpus and/or parallel corpus for other studies⁶. Materials are pre-processed, individualised and aligned. The corpus includes c. 1.1 million running words of source documents and approx. the same number for the translated texts. For the purpose of conducting this analysis the corpus was not searched in its entirety and the author processed only the source text data.

With regard to the corpus design principles which hold true both for the corpus in its final shape as a translation/parallel corpus and in its source-texts-only variant exploited for the analysis in question the following comments need to be made. CorpCourt accounts for the principle of homogeneity in that the documents belong to the same context, domain, setting, purpose. The author used systemic search criteria which means that the search criteria were defined by spatiotemporal factors, that is (i) the year of entry in the register and (ii) the category of registration proceedings. Specifically, CorpCourt comprises an exhaustive set of documents for the year 2017 as found on court files for the registration of a branch of a foreign company in 1 out of 21 divisions of the National Court Register in Poland. The principle of exhaustive search and the fact of having searched an adequate proportion of the court localisations (2 out of 21 units of National Court Register) and systemic pre-definition of court files to be searched testifies to the reliability of the results.

The operationalization of the hypothesis involves calculating the frequency of selected verbal structures and comparing the results against the variables picturing the situation for genre-related, diatopic and diachronic perspective. Verbal structures were selected since they are claimed to constitute an important component of specialised languages in general and their distribution and use is said to be genre specific (See: Gotti, 2005; Williams, 2005). The nine types of verbal structures were selected on the principle of being basic forms and used extensively in the texts making up the corpus, which was established via visual pre-analysis of the corpus. For the sake of clarity, the term verbal structures is used with reference to the main categories covered by analysis and visualised on the radarplots (Figures 1–5) with abbreviated labels. Whenever the author refers to the forms or types then specific categories of verbal structures are meant.

This study was developed through the following stages: (i) lexical data processing was performed using Sketch Engine software; (ii) statistical computing was performed with the R tool; (iii) the data were normalized; and (iv) the frequencies of the features analysed were calculated for individual genres (predefined functionally) and with the variable of country and year of document publication. The texts making up the corpus involve the following categories: CONFIRMATION OF REGISTRATION, COMPANY EXTRACTS, FOUNDATION ACTS, DECLARATIONS OF WILL, FINANCIAL DOCUMENTATION, REPORTS, AUTHENTIFICATIONS, AUTHORISATIONS, VERIFICATIONS, RESOLUTIONS and MISCELLANEOUS. For calculating the tendencies with regard to the variable of year, country of origin the metadata were standardised to the following broader categories: (i) country: US, UK, EU, AS, CA and (ii) year: <2000, 2001–2010, >2010.

⁶ The author uses the term parallel corpus as stipulated, for example by (Ebeling, 1998; Więclawska, 2018; Więclawska, 2019).

The steps in the R analysis involved plotting radar-plots data pre-processing. PCA analyses were performed to see if there are any clear clusters. Unfortunately in each case the proportion of variance carried out by the two first principal components was low.

3. RESULTS

Plotting the frequency data allowed the author to arrive at the distribution schemes of the verbal structures covered by the analysis in the context of the variable of genre, country of origin and time of publication of the documents making up the corpus. The verbal structures scrutinised include the following, starting from the top, reading the Figure in the 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*. Figure 1 visualises the data for the genre referred to as FINANCIAL DOCUMENTATION but it is held as a fairly close model, representing the average data distribution for all the texts in the corpus.

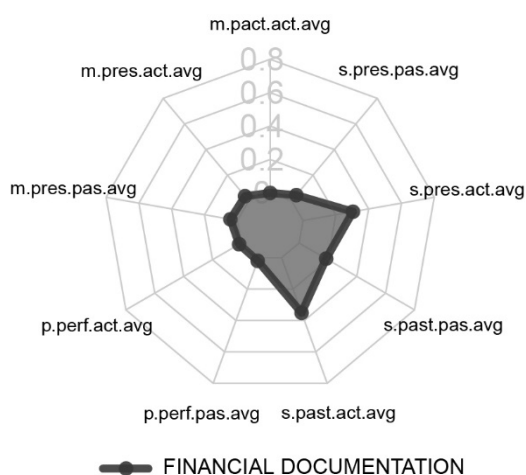


Figure 1. Model frequency distribution of verbal structures for the CorpCourt texts

It is evidenced that majority of the verbal structures present equal participation in the usage and three of them show themselves as salient values. The latter class includes simple present active forms, simple past passive forms and simple past active forms. The extensive use of basic grammatical tenses, as is the case here shows the texts making up the corpus present high level of generality (See: Gotti, 2005).

The foregoing discussion is not intended to refer to all the aspects of the material compared. The author refers only to selected patterns within a verbal structure which is shown to be most prominent statistically and/or relevant as a rhetorical tool of text stylistics. The discussion proceeds from the identification of the verbal structure which causes stylistic variability in the context of the variable of genre, place of origin and time of document publication, and discussing selected aspects of this verbal structure. Notably, even when

comparison of two different distribution patterns yields a few differences, the author limits the discussion to one instance because of limitation of space and the scope of the data processed.

The general frequency distribution is visualised in the Figures in the text but the author uses additional data extracted from Sketch Engine performing further queries with the functionality of filtering and sorting the material so that the operation of the three variables is accounted for. Importantly, the truly comparative component of the discussion is limited to identifying the distinctive linguistic features, that is the extensive use of given verbal structures. The data discussed further in relation to these features refer exclusively to various aspects of the verbal structure, as found distinctive in the context of a given variable. The discussion at this stage does not focus on the aspects that constitute the material causing distinctions. Specifically, the data related to the tendencies evidenced on the relevant Figures below and discussed, refer consistently to the specific categories within the salient verbal structures (types/forms), their semantic profile and – if possible – some tendencies with regard to the closest collocates.

Figure 2 below presents the data that are illustrative of a case of stylistic distinction of two genres. The data were extracted by applying the querying formula: [tag="VVP|VVZ|VBZ|VBP"][[tag!="VVN|VBN"]], which involves the elimination of the so called noise and unwanted structures. The most significant differences between the genre-related contexts compared seems to be the case of simple present active forms which show to be dominant in the texts related conceptually to the category COMPANY EXTRACTS when juxtaposed with the sub-corpus of the texts affiliated to the conceptual category FINANCIAL DOCUMENTATION.

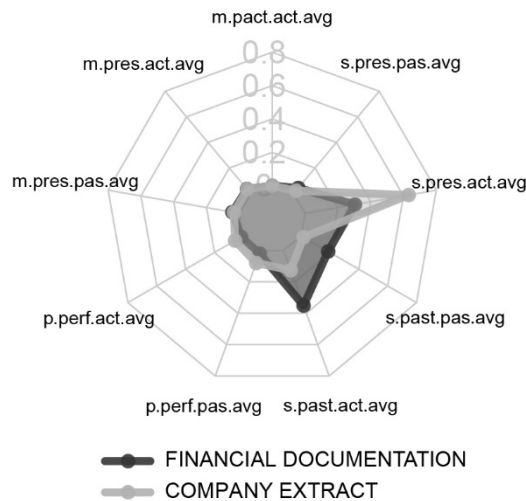


Figure 2. Comparative account of the frequency distribution of verbal structures in the category FINANCIAL DOCUMENTATION and COMPANY EXTRACTS – salient simple present active forms

The scores for the two genres are 631 hits (561,29 per million) and 592 (627,48 per million) in favour of COMPANY EXTRACTS. More detailed processing of the material allows us to discern various types of verbs which fit in the category of verbal structures labelled as simple present active. The top agent on the list here is the third person singular form of 'be', both in negations or affirmative sentences. These forms account for more than half of all the instances of simple present active (229 hits/203,7 per million out of the total of 631 hits/561,29 per million). The high frequency of the third person singular forms of 'be' often followed by predicative adjectives acting as subject complements or predicative adjuncts acting as complements of the copular 'be' can be attributed to the preference for simplicity. Sophisticated forms are avoided and basic verbs with a wide scope of application are used instead. This principle shows to be more in force for COMPANY EXTRACTS which is justifiable on the grounds of low complexity of the contents and purely, standardised communicative function of the said texts. Also, there are numerous instances of third person verb forms followed by the pro-form 'hereby' and these are considered as markers of performative forms, which remains one of the most common characteristics of legal language, being exceptionally distinctive for the genre of COMPANY EXTRACTS, as emerges from the data.

Furthermore, some of the context in which simple present active structures are used in the texts fitting the category of COMPANY EXTRACTS confirm the assumption made about grammar of specialised languages saying that some verb forms are used not according to the time axis of the events in question but as a tool to fulfil some rhetorical function. The case in point seems to be consistent use of third person singular 'be' followed by infinitive to express command or contingent future.

Generic distinctions with regard to the verbal structures are also visible when we juxtapose the texts classified as FOUNDATION ACTS and COMPANY EXTRACTS. Figure 3 shows that *modals with present reference, followed by the active infinitive* stand out when compared with the documents categorised as COMPANY EXTRACTS.

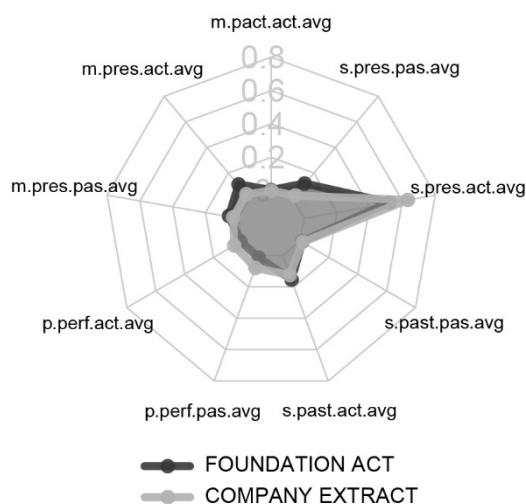


Figure 3. Comparative account of the frequency distribution of verbal structures in the category FOUNDATION ACT and COMPANY EXTRACTS – salient modals

The extraction of the relevant material was performed by operation of the query: [tag="MD"&word!="will"] [tag="RB"]{0,1} ([tag="VV"] [tag="VH"] [tag="VB"] [tag!="VVN"]). This allowed for elimination of the noise and unwanted structures and – after having filtered the material for the salient genre – provided us with frequency data for the documents classified as FOUNDATION ACTS. *Modals with present reference followed by active infinitive* account for 4,402 hits (3,916.66 per million) as compared to COMPANY EXTRACTS which the score of 146 hits (129.87 per million).

There are a few comments to be made at this point. Firstly, closer examination of the categories of the said modals here points to the high position of *shall*, which is in line with the empirically evidenced assumptions about the increased use of deontic *shall* in legal texts. It is evidenced that this tendency is stronger for FOUNDATION ACTS. Secondly, the unequal distribution of the frequencies in this respect points to high correlation between the communicative function of a text and the use of these verbal structures. FOUNDATION ACTS are found to include more rhetorical tools that serve as a vehicle for imposing/stating rights. When used in its deontic capacity, *shall* is the main tool to express obligation, which is of significant application in the texts types scoring higher results in frequency of usage. Thirdly, the frequency data for the individual modal verbs used in the context referred to as modals with present reference followed by active infinitive show huge disproportions in the distribution, which may be considered to be the peculiarity of the genre in question. As mentioned, the list of modals with present reference, followed by active infinitive is topped by *may*, scoring 2,077 hits (1,847.53 per million). *May* is followed by *shall* with a score of 1,983 (1,763.92 per million). The remaining categories such as, for example, *must*, *would*, *can* and *need* fall well behind in terms of scores. This points to the highly repetitive structure of legal texts categorised as FOUNDATION ACTS, also confirmed by the repertoire of the lexical items constituting the immediate, first-to-the-right context. Again the disproportion here is significant in that the first 5 hits show formula scoring 100 hits or more with *shall not apply* as a leader (226 hits (201.03 per million), while the number of occurrences for the other modals constitutes less than one third of the top scores.

3.1. Diatopic perspective

With regard to the diatopic perspective, potential variation in the distribution of the lexical features in general is assumed to be operative on the grounds of differences in the legal systems to which the said texts belong. The influence of this factor on the stylistic structure of the texts remains unquestionable and it is materialised, among others, by virtue of the principle of intertextuality which obligates the entities drafting legal texts to use prefabricated, constantly recurring phrases from the statutory instruments. If the latter differ cross-nationally the texts classified as secondary genres will display stylistic differences as well.

Analysis of the material from the diatopic perspective is promising and revealing in view of the range of foreign countries and thus system represented by the texts making up the corpus. These are not only instances of the UK or US varieties but also texts from other Anglophone countries, where English has the status of first official language, second official language or is the language of corporate communication in the capacity of the lingua franca.

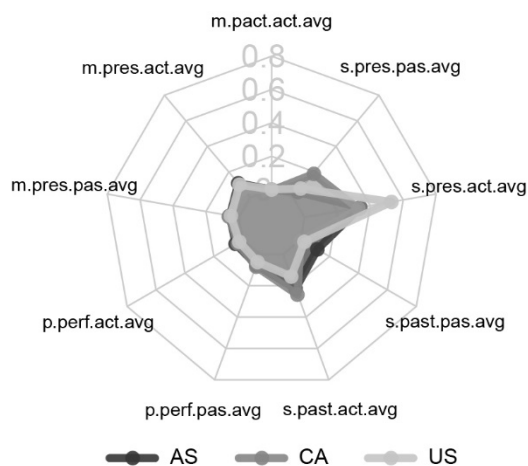


Figure 4. Diatopic account of the frequency distribution with regard to verbal structures – salient simple present active forms

The data visualised in Figure 4 evidence a few salient areas. One of them is the relatively increased frequency of simple present active verbal structures in the texts originating in the US. The technical aspect referring to the query formula with respect to this verbal structure was already referred to.

Referring to the US background, the overall score for the US with regard to the said verbal structures is 1,110 hits (987.32 per million). This number is yielded by Sketch Engine after the data were filtered according to the possible noise factor (hits followed by past participle form, which would constitute cases of passive voice, were eliminated) and country of origin of the document publication.

The linguistic material that proves to demonstrate the increased frequency on the radar-plot (Figure 4) can be split into the categories corresponding to the types of verb forms. The score in point is as follows: third person singular simple present tense lexical verbs other than *be* top the list with 453 hits (402.95 per million); this category is followed by the third person singular simple present verb *be* scoring 373 hits (331.79 per million); with two further, minor categories with decisively decreased scores, and these are non-third person singular simple present lexical verbs other than *be* evidenced with 181 hits (161.00 per million) and non-third person simple present forms of the verb *be* with 103 hits (91.62 per million). Three comments may be made at this point. Such a frequency distribution shows preference for the third person forms in the said texts, whether these are lexical verbs or simple present forms of *be*. Third person form is impersonal and it is often exploited in performative formula, which defines legal discourse in general to some point, and proves to be a specifically strong peculiarity of the simple present active forms in the texts originated in the US. Moreover, the high position of *be* in simple present and *do/does*, both found in the top 10 word forms testifies to the evident preference for the use of verbs of wide application. Finally, the semantic profile of the lexical verbs proves to be highly schematic. The top 10 lemmas are dominated by the verbs that operate as authority markers, either by the act of legitimising acts in law (*approves*, *certifies*, *bears* [the signature]) or by the act of exercising rights (*adopts*, *resolves*, *requires*).

3.2. Diachronic perspective

The diachronic perspective visualized on Figure 5 below was supposed to examine whether legal grammar is subject to variation in time, with focus on the scrutinised verbal structures and – if ‘yes’ – which verbal structures are affected. The hypothesis related to variation in time rests on the positive empirical evidence on the time-related variation in legal terminology (See: Peruzzo, 2017), including the short-period perspective. The time span variable applied in this study is visualised as the three temporal categories according to which the texts making up the corpus were classified. The period taken into consideration is determined by the date of publication of the said documents and it covers approximately 20 years. There are literally a few documents dated in the 1960’s but these make a statistically insignificant sample.

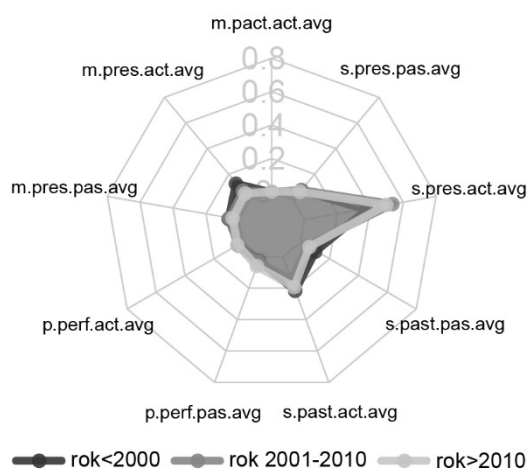


Figure 5. Short-term diachronic account of the frequency distribution with regard to verbal structures

As evidenced in Figure 5 no significant variation is noted with regard to the use of the verbal structures under analysis. None of the lexical features analysed shows as distinctively salient in any of the specific time slot. The distribution patterns corresponding to the individual periods in principle are overlapping, with simple present active and simple past active forms in a considerable majority, modal present active forms fairly salient as regards the frequency load and the other verbal structures staying at the level of centre-close circles on the radar-plot.

This finding seems to confirm the general assumption about the slow dynamics of legal language (See: Lemmens, 2011; quoted after Peruzzo, 2017) even in longer time perspectives. It may also be the case that verbal structures prove to be more change resistant in the diachronic perspective that other areas of legal language, terminology being an example of the latter (simplification, proliferation of new terms referring to one concept etc.).

4. CONCLUSIONS

The discussion presents results that foreground the features of legal language against the background of specialised languages in general. The added value of the findings consists in

presenting results related to the so-called secondary genres of legal communication, and there are few studies that do that, especially as regards authentic texts. Specifically, the data complements the findings gathered in the research on prescriptive texts, mainly EU texts. Additionally, the research addresses the need for more field- and context-specific research in that it relates to the domain of corporate law, focusing on the company registration discourse thus enabling future comparative studies setting legal stylistics in general against domain specific stylistics (company registration discourse).

Some of the findings confirm the well-established assumptions related to the specificity of specialised communication on the ground of legal communication, and specifically distinctive features of verbal structures. Here belong (i) the prevalence of present indicative (See: Gotti, 2005) (ii) the correlation between the frequency scores of specific data and text types with regard to their communicative purpose (See: Gotti, 2005) and (iii) the fact that it is often the rhetorical functions that determine the verb forms and not the time axis (See: Gotti, 2005; Lackstron, Selinker, & Trimble, 1973). The discussion also unveils some correlation or lack of correlation between the said verbal structures and some sociolinguistic factors presenting the lingua franca dimension of English.

It remains the author's hope that, apart from being a research exercise which – hopefully – complements the bulk of research on English legal language, the research provides valuable and reliable material for didactics of specialised languages. Teachers receive a confirmed repertoire of grammatical structures that recur through the said area of legal communication, which allows them to authenticate the teaching process and profile the teaching curricula in a rational and market-related way. What is of significance, the materials serve as a good source for compiling context-specific and usage-based functional equivalents for a number of rhetorical structures and routine formula to be used in translation, both with regard to teaching translation and practising translation.

It needs to be admitted that in some respects, even within the realm of verbal structures, the analysis did not exploit the problem area in total. For more complete results the research could be extended by covering new disciplinary domains of legal communication and – with regard to the structures analysed – it could cover other verbal structures, including the non-finite constructions. It is to be verified whether more in-depth qualitative analysis could disclose some functionally significant yet frequency-wise less prominent structures which remain distinctive markers of the registration discourse stylistics. This can hopefully be achieved by more fine-grained techniques of querying the corpus, involving various combinations of metadata filtering and by applying processing some the material for sociolinguistic data with regard to the individual genres.

REFERENCES

- Baker, M. (2004). *A corpus-based view of similarity and difference in translation*. "International Journal of Corpus Linguistics", No. 9.
- Bhatia, V. (2006). Legal genres [In:] Brown, K., ed., *Encyclopedia of Language and Linguistics*. Oxford: Elsevier.
- Bhatia, V.K. (2017). *Critical Genre Analysis. Investigating interdiscursive performance in professional settings*. New York: Routledge.
- Biel, Ł. (2014). *Lost in the Eurofog: the Textual Fit of Translated Law*. Peter Lang.
- Ebeling, J. (1998). *Contrastive linguistics, translation, and parallel corpora*. "Meta", No. 43(4). DOI: 10.7202/002692ar.

- Gotti, M. (2005). *Investigating Specialised Discourse*. Bern: Peter Lang.
- House J. (2013). *English as a lingua franca and translation*. "The Interpreter and Translator Trainer", No. 7(2).
- Lackstron, J., Selinker, L. & Trimble, L. (1973). *Technical Rhetorical Principles and Grammatical Choice*. "TESOL Quarterly", 7.
- Legallois, D., Charnois, T. & Meri Larjavaara. (2018). *The balance between quantitative and qualitative literary stylistics: how the method of "motifs" can help*. [In:] Legallois, D., Charnois, T. & Larjavaara, M., eds., *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*. Berlin: De Gruyter Mouton.
- Lemmens, K. (2011). *The slow dynamics of legal language: Festina lente? "Terminology"*, No. 17(1).
- Longerée, D. & Mellet, S. (2018). *Towards a topological grammar of genres and styles: a way to combine paradigmatic quantitative analysis with a syntagmatic approach* [In:] Legallois, D., Charnois, T. & Larjavaara, M., eds., *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*. Berlin: De Gruyter Mouton.
- Peruzzo, K. (2017). *Legal system: an additional variable in the analysis of short-term diachronic evolution of legal terminology*. "International Journal of Legal Discourse", No. 2(2).
- Pontrandolfo, G. (2015). *Investigating judicial phraseology with COSPE: A contrastive corpus-based study* [In:] Fantinuoli, C. & Zanettin F., eds., *New Directions in Corpus-based Translation Studies*. Berlin: Language Science Press.
- Więclawska, E. (2018). *Towards the establishment of genre distinctions of company legal registration proceedings*. "Komunikacja Specjalistyczna", No. 16.
- Więclawska, E. (2019). *Field-specific conventions in the translation of company law documentation for court proceedings*. "Studies in Logic, Grammar and Rhetoric. Studies in Logic, Grammar and Rhetoric", No. 58(71).
- Williams, C. (2005). *Tradition and Change in Legal English. Vernal Constructions in Prescriptive Texts*. Bern: Peter Lang.

DOI: 10.7862/rz.2019.hss.48

The text was submitted to the editorial office: September 2019.

The text was accepted for publication: December 2019.

