dr Edyta Więcławska
Uniwersytet Rzeszowski
Instytut Neofilologii, Katedra Filologii Angielskiej
tel. (17) 872 12 20
e-mail: edytawieclawska@poczta.fm
ORCID ID: https://orcid.org/0000-0003-0798-1940

DISCRETE UNITS AS MARKERS
OF ENGLISH/POLISH CONTRASTS
IN COMPANY REGISTRATION DISCOURSE

ABSTRACT

The paper addresses the issue of the complexity of legal communication in an interlingual perspective. The analysis fits in the paradigmatic approach to contrastive studies, where the distribution of discrete units is presented in quantitative terms. The cross-linguistic, computational account of the distribution of selected discrete units across company registration documentation shows systemic distinctions. It discloses recurrent, symmetrical and asymmetrical patterns which result from language system-inherent distinctions and/or the operation of translation universals. The strong point of the research lies in addressing legal communication within the realm of secondary genres, which – for practical reasons – are underrepresented in jurilinguistic studies. The study is based on a custom-designed, parallel corpus comprised of authentic materials.

Key words: legal language, discrete units, interlinguistic contrasts, legal genre

ABSTRAKT

JEDNOSTKI DYSKRETNE JAKO WYZNACZNIKI ANGIELSKO/POLSKICH KONTRASTÓW W DYSKURSIE REJESTRACJI SPÓŁEK PRAWA HANDLOWEGO

Artykuł dotyczy złożoności komunikacji prawniczej w kontekście różnic pomiędzy językiem angielskim i polskim. Analiza należy do nurtu badań kontrastywnych polegających na badaniu ilościowego rozkładu jednostek dyskretnych. Przedstawiony rozkład wybranych jednostek dyskretnych w tekstach dotyczących rejestracji spółek prawa handlowego wykazuje systemowe różnice. Analiza potwierdza istnienie powtarzalnych, symetrycznych i asymetrycznych schematów, które wynikają różnic pomiędzy językami

i/lub działania uniwersaliów tłumaczeniowych. Mocną stroną tego badania jest fakt, że zostało przeprowadzone ma tekstach prawniczych, które – z przyczyn praktycznych – rzadziej są przedmiotem analiz. Badanie zostało przeprowadzone na autorskim, paralelnym korpusie, składającym się z autentycznych materiałów.

Słowa kluczowe: język prawa, jednostki dyskretne, kontrasty interlingwalne, gatunki tekstów prawa

1. Introduction

Legal communication has been the subject of multifaceted contrastive analyses which were conducted from both the intralingual and interlingual perspectives. The first type of contrastive analyses focuses largely on register- and mode-related variations[1], translated vs. non-translated distinctions[2] or shifts in diachronic perspective[3], while the latter naturally brings to the fore the cross-linguistic contrasts[4]. This study addresses both these areas in that ultimately it attempts to identify some aspects of the English-Polish contrasts in the field of legal language, but it also captures the intralinguistic contrasts from the perspective of variation in the source English language texts.

Apart from the above-specified aspects of contrastive studies in the area of legal language that are addressed by this study, a few other linguistic traditions can be mentioned on which this study draws: corpus methodology[5],

---

[1]   See, for example, S. Goźdź-Roszkowski, *Patterns in Linguistic Variation in American Legal English*, Frankfurt am Main 2011.

[2]   See, for example, Ł. Biel, *Lost in the Eurofog: the Textual Fit of Translated Law*, Frankfurt am Main 2014.

[3]   See, for example, K. Peruzzo, *Legal system: an additional variable in the analysis of short-term diachronic evolution of legal terminology*, "International Journal of Legal Discourse" 2017, no. 2(2), pp. 291–313.

[4]   See, for example, R. Rabadán, *Translating the "predictive" and "hypothetical" meanings English/ Spanish*, "Meta" 2007, no. 52(3), pp. 484–502; R. Salkie, *Legal phraseology in contrast. The fact that and its German counterparts*, [in:] *Phraseology in Legal and Institutional Settings. A Corpus-based Interdisciplinary Perspective*, S. Goźdź Roszkowski, G. Pontrandolfo (eds.), London–New York 2018, pp. 126–143; N. Szymor, *Behavioural profiling in translation sudies*, "Trans-Kom" 2015, no. 8(2), pp. 483–498; F.J. Vigier, M. del Mar Sánchez Ramos, *Using parallel corpora to study the translation of legal system-bound terms: The case of names of English and Spanish courts*, [in:] *Computational and Corpus-based Phraseology. Second International Conference, Europhras 2017 London UK, November 13-14, 2017 Proceedings*, Cham 2017, pp. 260–273.

[5]   See, for example, G. Vanderbauwhede, P. Desmet, P. Lauwers, *The shifting of the demonstrative determiner in French and Dutch in parallel corpora: From translation mechanisms to structural differences*, "Meta" 2011, no. 56(2), pp. 443–464; *Podstawy językoznawstwa korpusowego*, B. Lewandowska-Tomaszczyk (ed.), Łodź 2005.

discrete units-based analyses[6], and sociolinguistics in specialised communication[7].

The study makes up part of a larger project on English/Polish contrasts in legal language in the said domain, where quantitatively salient grammatical categories are examined for more detailed cross-linguistic distinctions in the function-based and translation-related perspectives[8]. At the stage of the project covered by this discussion the study aims at establishing what the English/Polish contrasts in legal language are, whether they are domain-specific, and whether they contextually vary in intensity. The primary aim is also to identify agents/ the excess categories reflecting the cross-linguistic contrasts, with the aim of customising the forthcoming deeper qualitative analysis in the sense of identifying the discourse/context specific contrasts only.

The hypothesis was formulated that certain categories of discrete units have cross-linguistically (English/Polish) distinct frequency distribution throughout legal language and the scale of these contrasts may be genre-dependent. Furthermore, irrespective of their intensity, the contrasts allow themselves to be systemically accounted for on the grounds of the translation process and systemic distinctions between the languages.

To test the hypothesis posed, the author formulated three research questions:

1. What is the frequency distribution for the individual grammatical categories/word classes/discrete units covered by the analysis, in the whole corpus and in specific genres?
2. Is the salience of specific grammatical categories more marked in some genres?
3. What are the tentative findings with regard to the factors triggering the cross-linguistic distinctions in question?
4. What are the interlingual correspondence patterns that emerge from the quantitative and qualitative analysis?

---

[6] See, for example, D. Longerée, S. Mellet, *Towards a topological grammar of genres and styles: a way to combine paradigmatic quantitative analysis with a syntagmatic approach*, [in:] *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, D. Legallois, T. Charnois, M. Larjavaara (eds.), Berlin 2018, pp. 140–163.

[7] R. Dodsworth, *Speech communities, social networks, and communities of practice*, [in:] *Research Methods in Sociolinguistics. A Practical Guide*, J. Holmes, K. Hazen (eds.) Hoboken 2014, pp. 262–275.

[8] The discussion here presents an extension of the findings described in E. Więcławska, *English/ Polish contrasts in legal language from the usage-based perspective*, *Translata III Tagesband*, 2020, to be published. The paper referred to initiated the analysis based on discrete units-based cross linguistic distinctions in legal language. The data presented here extended the initial findings by covering additional 5-word classes, all the genres identified in the corpus, extending the qualitative analysis and performing additional statistical significance analyses.

2. Methodology

The methodology involves processing the custom-designed corpus (hereinafter referred to as *CorpCourt*), which includes c. 1.1 million running words of source documents in English and approximately the same number for translated texts in Polish[9]. Although criticised in some studies, the methodology based on discrete units is assumed to be reliable here on the grounds of the significant functional and contextual parallelism of the sub-corpora in the two languages, thus their high degree of thematic and grammatical homogeneity and elements of parallel qualitative analysis based on the aligned units. These aspects are assumed to legitimise the findings. It also needs to be stated here that the discrete-units methodology makes it possible to achieve the background aim of the study consisting in acquiring contextually exhaustive data (company registration discourse) about structurally distinctive areas in the two languages, and thus adopt an a priori/agent-first approach towards tracing cross-linguistic contrasts. Distinctions in the frequency distribution of discrete units allow us to identify the focal and primary points responsible for potential discursive/functional/ stylistic contrasts, in order to identify the agents of functional/stylistic contrasts to be studied further. It is assumed that the asymmetrical frequency distribution of the specific categories is indicative of the cross-linguistic contrasts in the grammatical/stylistic structure (the same functions expressed with different word classes or shifts caused by the process of translation), the cross-linguistic distinctions in frequency are not shared by legal language in general but are context specific (company registration discourse) and – thus – a reliable qualitative study should start with quantifying context-specific discourse units.

Referring to the status of the language material covered by the analysis, the study complements the existing research with findings based on bilingual (translation corpus), authentic and highly specialised data[10]. Additionally, the language material compared covers only texts classified as secondary genres

---

9   The *CorpCourt* in its interim composition has been exploited for analyses addressing other areas of linguistic phenomena, including the intralinguistic and interlinguistic perspectives. See, for example, E. Więcławska, *Field-specific conventions in the translation of company law documentation for court proceedings*, "Studies in Logic, Grammar and Rhetoric" 2019, no. 58(71), pp. 221–243; E. Więcławska, *Language of Law: Court translation in the didactic perspective*, [in:] *Specialist Communication. Student Training and Market Demands*, E. Grygiel, A. Uberman (eds.), Rzeszów 2018, p.p 37–60; E. Więcławska, *Sociolinguistic and Grammatical Aspects of English Company Registration Discourse*, "Humanities and Social Sciences" 2019, vol. XXIV, no. 26(4).

10   F.J. Vigier, M. del Mar Sánchez Ramos, op. cit., pp. 4–5 emphasise 'Scarcity of legal, bilingual corpora', and they claim that '[…] such specialized ad-hoc corpora have to be compiled', adding that '[…] Specialised research into translation phenomena requires a corpus that can shed light on specific translation issues'.

related to a subfield of corporate legal language, that is company registration proceedings, and secondary genres are said to escape linguistic analyses.

The operationalization of the hypothesis involves calculating the frequency of seven grammatical categories acting as *tertium comparationis* (adjective, nouns, coordinating prepositions, subordinating prepositions, adverbs, pronouns, and prepositions) for the whole corpus and for the individual genres, and subsequently comparing the results against the independent variable of genre. This study was developed through the following stages: (i) lexical data processing was performed using Sketch Engine software and candidate terms were extracted; (ii) statistical computing was performed with the R tool; (iii) the data were normalized; (iv) the frequencies of the grammatical categories analysed were calculated for individual genres (predefined functionally), and (v) qualitative analysis covering the highest ranking candidate terms. The texts making up the corpus involve the following categories of genres: CONFIRMATION OF REGISTRATION, COMPANY EXTRACT, FOUNDATION ACT, DECLARATION OF WILL, FINANCIAL DOCUMENTATION, REPORT, AUTHENTIFICATION, AUTHORISATION, VERIFICATION, RESOLUTION, and MISCELLANEOUS.

The discussion starts with a presentation of the overall frequency distribution in contrast, and it proceeds to presenting quantitative and qualitative data for the three word classes that scored the highest results. At this stage the quantitative assessment was made on the basis of normalised frequencies, as visualised in Figure 1. The three sections of the discussion involve presenting the holistic data, that is statistics for the whole corpus, and a more detailed discussion of word class-specific data for a selected genre which is representative for the tendency of quantitative salience and/or qualitative profile. At the latter (genre-specific) stage the discussion involves a qualitative analysis examining the capacity of the relevant parallel data within the individual word class/grammatical category for (1) semantic equivalence (vertical dimension) and (2) corresponding frequency level (horizontal dimension). *Parallel data* here means the highest-ranking candidate terms, ranked above the threshold of the 5 top items on the frequency list. The level of semantic equivalence is measured by quantifying the number of pairs of equivalents above the said threshold, and the equivalence is established on the grounds of the frequency of the first definition evidenced in the legal English dictionaries consulted (*dictionary equivalence*). The stage corresponding to the horizontal approach consists in assessing the level of frequency level correspondence between the pairs of semantic equivalents identified in the vertical assessment. It is believed that such methodology may be effective in the highly homogeneous and thematically contiguous environment, as is constituted by our parallel, translation-based corpus. Fragmentary qualitative analysis based on the sample parallel data consisting in the examination of the level of semantic

and frequency-related contiguity between the top representatives of the 3 word classes that are cross-linguistically distinctive in terms of frequency distribution provides empirical ground for the identification of factors responsible for the cross-linguistic distinctions.

The discussion includes rough frequency data to support the arguments, but the radar plot calculations and related findings are based on normalised data. Statistical significance tests were performed for the three word classes only for the holistic data.

3. Discussion

The data covered seven word classes, which are, starting from the top-most in the clockwise direction: adjectives, subordinating conjunctions, coordinating conjunctions, pronouns, prepositions, nouns and adverbs. They have been statistically processed for both languages, and their distribution is presented visually as an overlapping pattern.

Figure 1 shows that (i) the frequency of the individual categories/word classes covered by the analysis varies in the corpus between the languages, and (ii) the rough distribution pattern for them displays analogical proportions for the two languages, testifying, for example, to a significantly larger share of nouns and prepositions over adverbs or subordinating conjunctions. The frequency data for the individual categories/word classes are as follows: (i) for Polish: adjectives (124,861/124,100.12 per million), subordinating conjunctions (7,353/7,484.86 per million), coordinating conjunctions (47,964/48,805.81 per million), pronouns (8,257/8,405.07 per million), prepositions (107,346/409,271.03 per million), nouns (333,368/229,346.26 per million), and adverbs (14,981/15,249.65 per million), and (ii) for English: adjectives (56,998/50,701.65 per million), subordinating conjunctions (25,075/22,304.67 per million), coordinating conjunctions (47,901/42,608.81 per million), pronouns (13,710/12,135.3 per million), prepositions (160,158/142,463.47 per million), nouns (347,114/308,309.7 per million), and adverbs (25,864/23,006.5 per million).

Figure 1 presents averaged data, without referring to the individual genres. The data constitute a reference scheme on the basis of which the quantitatively salient categories were selected for detailed discussion in the foregoing. These are: adjectives, prepositions and nouns.

Adjectives have attracted the attention of researchers of legal language most often in the contexts of their premodification, postmodification capacity, structures they make up a part of, and their role as phrasal constituents, including

Fig. 1. Distribution of seven word classes in company registration discourse – holistic account of the English/Polish contrasts.
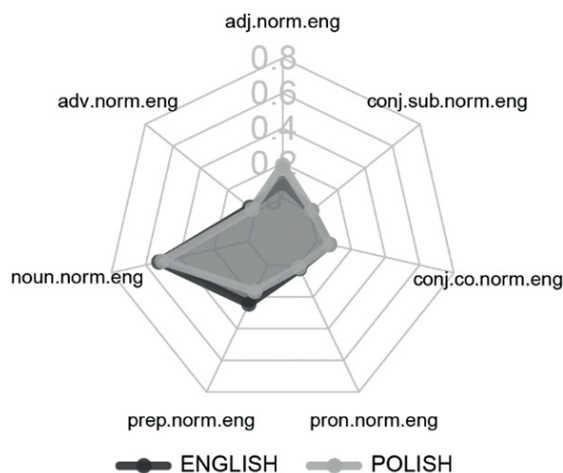
the fixed phrases perspective[11]. Although the findings for English are that their overall frequency in legal language is reported to be lower than in general language, even with the decreased averaged volume functions and levelled syntactic functions, they remain a distinctive feature of legal discourse[12]. What is of relevance for this analysis is that adjectives are extensively discussed in the literature of the subject in the context of interlingual communication, which is confirmed by a number of studies on cross-linguistic contrasts and correspondences in the said domain. Scholars point to the varied and complex status of adjectives in legal communication[13] which – either indirectly or directly – may constitute a guideline for the legal interpretation of the related structures and – if we are participants of interlingual communication – for translation[14]. The quantitative salience of the adjectives compared to other word classes seems to confirm the

---

[11]   C. Williams, *Tradition and Change in Legal English. Verbal Constructions in Prescriptive Texts*, Bern 2005, pp. 73–74.

[12]   M. Bázlik, P. Ambrus, *A Grammar of Legal English*, Bratislava 2008.

[13]   For more on the status of adjectives as derivatives of, or parts of, normative terms see, for example, A.L. Kjœr, *Phrasemes in legal texts*, [in:] *Ein Internationales Handbuch der Zeitgenössischen Forschung/Phraseology. An International Handbook of Contemporary Research*, Berlin – New York 2007, pp. 506–515; E. Wiesmann, *Rechtsübersetzung und Hilfsmittel zur Translation*, Tübingen 2004, pp. 360–361; E. Więcławska, *Extrapolating adjective distribution in Polish legal language*, [in:] *Specialist Languages in Use and Translation*, Rzeszów 2014, pp. 71–82.

[14]   E. Więcławska, *Extrapolating…*, op. cit.

assumptions voiced in the literature of the subject, and the English/Polish distinction makes it an interesting issue to investigate in more detail.

The overall data for adjectives are as follows: English scored 56,999/50,701.65 per million compared to Polish, registered at the level of 124,861/127,200.12 per million. The distinction proves to be statistically significant with the p value calculated from the normalised value 'per million' at the level of <0.001.

Taking account of the overall data and the data related to the individual genres, we can conclude that the frequency distribution of adjectives, which is consistent throughout the genres in the corpus here, may be said to confirm the specific position of adjectives throughout legal language, determined by their well-established status, including their role as components of various categories of norm-conditioned word combinations. Adjectives are a category that is resistant to generic conventions in legal communication and not vulnerable to individual stylistic preferences. It needs to be said at this point that the scale of the disproportion between the contrasted languages in this respect is so significant and consistent throughout the corpus that it must be a result of the systemic operation of differences in grammar and/or adopted translation techniques.

Figure 2 visualizes the quantitative salience of Polish in the genre referred to as FOUNDATION ACTS. For the category FOUNDATION ACTS, selected for more detailed analysis, adjectives score 61,769/82,876.6 per million in the Polish subcorpus and 30,717/27,323.22 per million in the English subcorpus. This tendency is noted for the other genres contrasted. The scale of the contrast is (i) relatively stable for all the genres represented in the corpus, and (ii) distinctly higher than for other grammatical categories contrasted.

The five top-most candidate terms qualified for detailed study are: *general, payable, present, ordinary, relevant* and – in the Polish subcorpus – *niniejszy, dowolny, walny, dany, żaden*. Vertical assessment of the 5 highest ranking candidate terms points to the limited cross-linguistical symmetry with regard to their semantic structure. The only relevant correspondences that were identified from among the 5 highest ranking candidate terms is *general*, corresponding to *walny*, and *relevant*, paired with *dany*[15]. The other 4 out of the 5 English terms found their matching, dictionary equivalents either slightly below the threshold, as with *ordinary*, held as equivalent with *zwykły*, placed in 6th position or – much lower on the list – as is the case with *present*, paired with *obecny*, and *payable*, matched with *płatny*.

---

[15]   *Relevant* was matched with two other dictionary equivalents which were ranked lower in frequency. For the purpose of this comparison, the highest ranking *dany* was chosen.

Fig. 2. Distribution of adjectives in company registration discourse – English/ Polish contrasts.

FOUNDATION ACT

ENGLISH POLISH
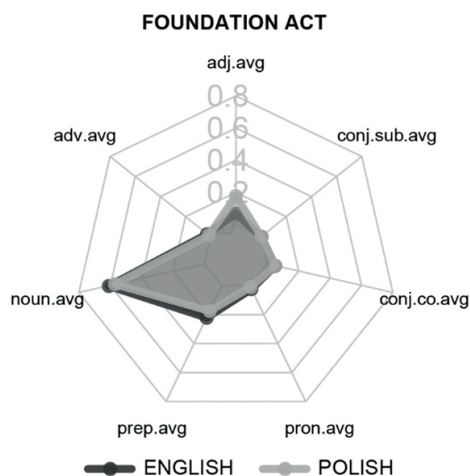
Contrasting the data from the horizontal perspective in the scope set for this level of analysis, that is juxtaposing the frequencies of the pairs that are semantic equivalents, does not provide us with unequivocal findings confirming cross-linguistically asymmetrical patterns. The picture here is largely varied and not consistent throughout the set of the 5 terms compared. Namely, noting the significant dominance of Polish at the level of total score for adjectives, at the level of individual pairs identified in point, Polish scores lower (*present, payable*), higher (*relevant, ordinary*) or almost equally (*general/walny*). Starting from the last category, the relevant cross-linguistic frequency data are as follows: *general* (1,017/904.64 per million) vs. *walny* (1,008/1,026.08 per million); *relevant* (427/379.82 per million) vs. *dany* (794/808.24 per million) and *ordinary* (463/411.85 per million) vs. *zwykły* (691/703.39 per million); *present* (472/419.85 per million) vs. *obecny* (442/449.93 per million), and *payable* (482/428.75 per million) vs. *płatny* (139/141.49 per million).

The cross-linguistic frequency distribution scheme with its two patterns: symmetry and asymmetry can be accounted for by (i) systemic differences between languages, and (ii) translation-related operations. Referring to the first aspect, a random check confirmed that the hiatus in frequency distribution between the corresponding dictionary equivalents (asymmetry) and the relatively low number of dictionary equivalents above the threshold set for the study (low level of symmetry) is triggered by the distinct grammatical affiliation of items materializing equivalent discourse functions. To explain, *niniejszy* (1,484/1,510.61 per

million), ranked among the top five in Polish, proves to be largely a translation equivalent for *this,* which is obviously not tagged as an adjective in the corpus. Another example is *dowolny* (1,119/1,331.07 per million), corresponding to *any*, the latter not being included in the list of adjectives. The same holds true for *żaden* (699/711.54 per million) and its translation equivalent *no*.

In the context of translation shifts, the random parallel check allows us to identify two types of tendencies in translation performance. Namely, in some cases adjectives prove to be translated with phrases that do not include adjectives. For example, *payable* happens to be translated as *do zapłacenia*.

The proliferation of synonyms in translation also comes as a recurring pattern. Hence, for example, *relevant* (427/379.82 per million) vs. *dany* (794/808.24 per million) + *odpowiedni* (324/329.81 per million) + *stosowny* (262/277.70 per million). Such cases lead to a proliferation of forms and thus a lack of one-to-one, recurring correspondence.

Acknowledging the operation of such translation shifts leading to cross-linguistic asymmetry in terms of lack of cross-linguistic correspondence with regard to frequency distribution of dictionary-based equivalents among the highest ranking candidate terms, we need to note a consistent tendency in translation as regards the recurrence of the same equivalents, when it comes to a certain category of adjectives which may be assigned the status of normative terms or may be components of normative terms. An illustrative example here is *general* and *ordinary* as part of the word combinations *general assembly/shareholders meeting*, translated as *zgromadzenie walne* and *ordinary shares*, translated as *akcje zwykłe*, respectively. Consistency in the translation of these terms is reflected in the comparable cross-linguistic frequencies of the adjectival pairs: *ordinary – zwykły* and *general – walny*. Such a pattern in translation performance in this respect confirms recognition of the special status of some adjectives by translators and thus their recognisable **readability** potential[16], confirmed by the resistance of these items to translation-related distortions.

Another category of discrete units that appears distinctive when comparing the Polish and English data is the category of prepositions (Figure 3). Prepositions have been extensively studied in contrastive analyses capturing the intralingual and interlingual perspectives, and have earned solid empirical evidence with regard to both general[17] and legal language (Biel, 2015). Studies contrasting specifically the English/Polish frequency data with regard to legal communication

---

[16]   See, for example, S. Wolfer, *Comprehension and comprehensibility*, [in:] *Translation and Comprehensibility*, K. Maksymski, S. Gutermuth, S. Hansen-Schirra (eds.), Berlin 2015, pp. 33–53.

[17]   A. Tyler, V. Evans, *The Semantics of English Prepositions. Spatial Scenes, Embodied Meaning and Cognition*, Cambridge 2003.

boil down to comparing the level of markedness of the contrasts between legal and general language in these languages. These findings and related distribution patterns viewed from the perspective of discourse functions are also discussed in the context of the synthetic/analytic classification of languages. Scholars emphasise the related cross-linguistic distinctions on this ground and point to the dynamic character of this situation. Namely, Polish is said to be contrasted with English by its more marked salience of prepositions in legal language compared to general language, as opposed to English, and the 'increasing analyticity of Polish' with regard to prepositions is noted[18].

Figure 1 presents the relevant data for all the genres. Polish recorded prepositions with a raw frequency of 107,346/109,271.03 per million, falling behind English, with a score 160,158/142,463.47 per million. The statistical significance test calculated from the normalized values (per million) was positive, with a significance level of p<0.037.

The findings from this study evidence a lower score for prepositions in legal Polish than in legal English, and as such they stay in line with the rough paradigm of synthetic languages as compared to analytic languages.

The analysis of the relevant material for the individual genres allows us to get more detailed findings in point. The genre DECLARATION OF WILL was chosen for closer analysis (Figure 3).

Figure 3 shows that in our corpus the number of simple prepositions in Polish is lower than in English. In principle, the tendency for quantitative overrepresentation of prepositions in English is reflected in the individual genres. For example, for DECLARATION OF WILL (Figure 2) English dominates over Polish with the score 8,233 (7,324.4 per million) compared to Polish 6,154 (6,284.36 per million). Notably, the salience is less marked in the genres under discussion than for the whole corpus. This seems to show a tendency that stays in line with the assumptions voiced in the literature of the subject, and mentioned above, about the 'increasing analyticity of Polish'. It seems the hiatus typical for analytic/synthetic language classification seems to become increasingly blurred with respect to the distribution of prepositions in legal language.

The quantitative results should be treated as complementing the findings gathered so far in the context of the scope of the research and methodology. Patterns identified in secondary genres in the field company registration discourse, as the object of this analysis, together with what was established with regard to legislative genres (Biel 2015), constitute a holistic picture of legal stylistics. From the technical point of view, the parallel corpus methodology applied here

---

[18] Ł. Biel, *Phraseological profiles of legislative genres: Complex prepositions as a special case of legal phrasemes in EU law and national law*, "Fachsprache" 2015, no. 3–4, p. 151, 155.

Fig. 3. Distribution of prepositions in company registration discourse – English/Polish contrasts.

DECLARATION OF WILL
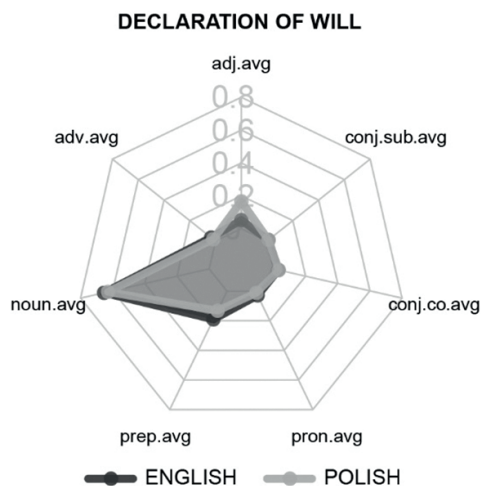
may be said to provide an insight from yet another angle into the grammatical aspects contrasted, providing a perspective that is inaccessible for the studies of comparable corpora.

A vertical account of the frequency data, that is taking account of the semantically-laden correspondences among the 5 top-most candidate terms, points to significant cross-linguistic symmetry in that the candidate terms overlap semantically. The candidate terms above the threshold of the top 5 hits include a relatively significant percentage of dictionary equivalents in the sense adopted in this paper, and these largely cover the most prototypical, spatial meanings. The leading categories in point here are carriers of locative, temporal meanings or items denoting the relation of inclusion. Specifically, three of the top five hits in point may be said to have dictionary equivalents in the corresponding Polish dataset. These are the pairs *in/w*, *to/do,* and *by/przez*. *For,* which is also above the threshold of the top 5 hits, does not depart much from the tendency in that it finds its semantically matching equivalent relatively close to the said threshold line. Finally, *of* shows itself to be a special case without a clear dictionary equivalent, because it is mainly used for the grammatical function of genitive, which in Polish is consistently expressed with an inflected structure[19].

---

[19]   Ibidem.

A comparative look at the highest-ranking candidate terms from the horizontal perspective, understood as comparison of the corresponding frequency data at the level of the pairs of the top 5 hits, points to traces of cross-linguistic asymmetry. Namely, the fact that the score of prepositions for English proves to be higher than for Polish does not always stay true for the individual scores for the pairs of equivalents compared. Specifically, the relevant frequency data for the said pairs are as follows: *to/do* (16,147/14,281.20 per million vs. 12,832/13,062.12 per million) and *by/przez* (11,766/10,466.07 per million vs. 7,223/7,352.53 per million) to the benefit of English. The reverse tendency is noted for the pair *in/w,* which evidences the dominance of Polish (21,676/19,281.20 per million vs. 32,087/32,662.41 per million).

To conclude, the phenomenon of English/Polish asymmetry in the frequency distribution of prepositions may somewhat tentatively be said to confirm certain structural differences between the languages and the operation of translation-related universals/shifts. Firstly, the contrasts noted are found to derive from language specific distinctions as regard grammatical structure. *Of* is a case in point here. Furthermore, overrepresentation of *w* compared to its dictionary English equivalent *in* shows that (i) *w* systemically takes on the function exercised by other English prepositions, and/or (ii) it makes up part of fixed phrases that are not translated literally from English, and thus the *in/w* proportions are distorted.

A random parallel check shows that from the point of view of the translation process/shifts, overrepresentation of prepositions in the general perspective may be accounted for by the process of simplification, whereby sophisticated variants of the same words, that is close synonyms, are rendered into Polish with repeatedly used simple equivalents. Cases in point are French or Latin words existing next to the English ones (e.g. *sine, cum, via, per*), translated as *bez, z, przez, przez,* respectively.

The symmetrical patterns identified in the analysis at the level of the highest-ranking prepositions were found to consist in significant overlap of semantic structure resting on the spatial relations. From the technical point of view, the parallel texts are found to largely rely on structures that both in English and Polish involve conceptualization via spatial relations.

Nouns are the third category that is found to display unequal proportions in our contrastive analysis, as visualized in Figure 5. This category encourages closer analysis also in view of its relatively high frequency salience in the two languages as compared to other word classes, and also due to its prominent status as a distinctive marker of legal stylistics in general. Scholars emphasize the

Fig. 5. Distribution of nouns in company registration discourse – English/ Polish contrasts.
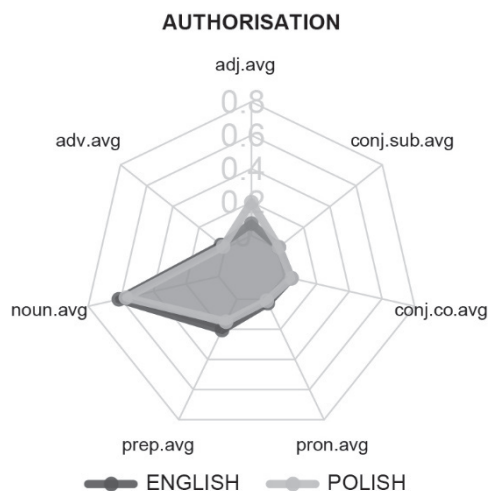
highly nominal style of legal discourse in general[20]. Nouns appear to be vulnerable agents in interlingual communication. Interlinguistic shifts are reported to occur 'at the level of the noun phrases' and they are 'mainly due to omission or addition of the noun phrase'[21].

Figure 1 shows the marginal domination of the English corpus over the Polish corpus with the following ratio: 346,603/308,309.7 per million for English as opposed to 333,368/339,346.26 per million for Polish. The p value was calculated at the level of 0.223 (p= 0.223), which points to the crosslinguistic distinctions not being statistically significant.

The gross statistics emerging from the analysis confirm the findings published in the literature of the subject with regard to both the increased salience of nouns as a grammatical category in the two languages and with regard to the tendency to distort the relevant marked contrasts in the cross-linguistic perspective. At the same time the distinction between the two languages proves to be marginal, which points to cross-linguistic symmetry with regard to nominal style. Let us take a closer look at the genre-specific frequency data.

---

[20] See, for example, in M. Gotti, *Investigating Specialised Discourse*, Bern 2005, p. 79; R. Hiltunen, *The grammar and structure of legal texts*, [in:] *Language and Law*, T. Peter, S. Lawrence (eds.), Oxford 2012, p. 47; A. Trosborg, *Rhetorical Strategies in Legal Language*, Tuebingen 1997, p. 14; C. Williams, 2005, op. cit., p. 37.

[21] See in G. Vanderbauwhede, P. Desmet, P. Lauwers, op. cit., p. 448.

For the genre AUTHORISATION the following quantitative data contrast the English and Polish subcorpus to the benefit of the English one: 14,284/12,705.88 per million contrasted with the Polish score of 12,584/12,809.67 per million. The tendency – although not universal for all the genres – proves to hold true for a few other genres represented in the corpus. These include texts classified as FOUNDATION ACT, REPORT and AUTHORISATION.

A vertical perspective proves cross-linguistic symmetrical patterns by identifying a proportionally high level of dictionary equivalents among the 5 items scoring the highest values. To illustrate, the representatives here include *company, date, director, share* and *name,* and four of them have their dictionary equivalents in the corresponding Polish subcorpus above the threshold of 5 (*spółka, data, dyrektor, akcja,* respectively) and the last one, i.e. *name*, has the matching dictionary equivalent in the first position below the adopted threshold (6th position).

A horizontal account of the 5 highest ranking candidate terms clearly confirms the already observed asymmetrical tendency 'higher vertically does not mean higher horizontally', which implies an increased density for the top representatives of the category of nouns in Polish (the frequency of the top hits in Polish is higher). Specifically, in spite of the higher gross/overall number of nouns in English, all 5 highest ranking candidate terms in English fall short of their Polish counterparts/equivalents in terms of frequency. The statistics are as follows: *company* (516/458.99 per million) vs. *spółka* (954/971.11 per million), *date* (367/326.45 per million) vs. *data* (388/294.96 per million), *director* (337/299.77 per million) vs. *dyrekt*or (360/366.46 per million), *share* (273/242.84 per million) vs. *akcja* (303/308.43 per million).

Qualitative analysis of the interlingually symmetrical patterns confirms the fact that the texts analysed feature domain specific terminology (commercial law), and this terminological macrostructure of the English texts is effectively transposed into Polish. Highly technical terms are translated into Polish with significant consistency. This may be held as proof of consistency in translation, which ensures maintenance of the genre's integrity and effective readability of the texts in translation.

Admitting the existing symmetry and – occasionally – the fact of marginality of cross-linguistic contrasts with regard to the frequency distribution of nouns in some respects, we cannot fail to notice clearly asymmetrical, cross-linguistic patterns in point. A random parallel search shows that the increased density of the 5 highest ranking candidate terms in Polish is due here to the operation of translation universals, the example in point being the case of explication or

convergence[22]. With regard to the first, pronouns in the source texts happen to be translated as the nouns to which these pronouns refer (e.g. *it/she – company*). Cases of convergence can be accounted for in two ways. Firstly, the intralingual, geographical variants of English provide manifold linguistic realizations of parallel concepts which prove to be 'flattened' and converge in Polish. A case in point is the UK *company* and US *firm* or *corporation,* translated markedly consistently as *spółka*. Secondly, there are cases of fine distinctions between specific terms that are not easily identified and accounted for in translation by non-lawyer translators. A case in point is the abstract/factual distinction, as in *incorporate* and *register*, the former related to the fact of gaining the legal capacity to act as a party to legal proceedings, and the latter related to the physical activity of entering a company into the register, both translated as *rejestrować*.

4. Conclusions

The findings encourage us (1) to draw some general conclusions from this stage of the research project on English/Polish contrasts, and (2) to identify the directions for further studies in this area.

To conclude, English and Polish legal language proves to have a fairly consistent, yet distinct frequency distribution with regard to the three word classes that prove to be over- or under-represented in Polish. These are the word classes that are commonly considered as characteristic for legal discourse in terms of frequency. In some cases these patterns of marked quantitative salience in a language/distinct frequency distribution are relatively genre-specific, but they are not subject to variation depending on the country of origin or year of document publication. The distinctiveness ratio/margin for overrepresentation for the individual genres varies in the case of some word classes. The markedness of the cross-linguistic contrasts is variably salient, depending on the word class (adjectives are significantly overrepresented in Polish, but the margin of overrepresentation for prepositions is insignificant).

Closer analysis of the material related to the individual word classes discloses cross-linguistic structural symmetry and asymmetry in the related areas. The concept of vertical comparison reveals patterns of semantic overlap among the items ranked highest on the frequency list. Asymmetry, understood as varied frequency of the dictionary equivalents, is also variable for the three word classes.

---

[22]   See, for example, Ł. Biel, *Lost in the Eurofog…*, op. cit.; E. Gumul, *Explicitation in Simultaneous Interpreting*, Katowice 2017.

The gross discrepancies in the cross-linguistic frequency distribution of the three word classes and the complementing comparison at the level of the highest ranking candidate terms can be accounted for in terms of the systemic differences between the languages and the operation of translation universals. The English and Polish languages have distinct grammatical structures with regard to the use of adjectives, nouns and prepositions. The differences are largely systemic in that the categories that are salient prove to recur throughout genres. Translation performance happens to be featured by the operation of simplification and explicitation in the said domains, and sometimes the comprehensibility/readability level on the part of translators is limited.

It remains to be hoped that the study is appreciated for gathering very domain-specific (company registration discourse) and authentic, contrastive data, based on solid statistical processing for a section of legal language that proves fairly resistant to legilinguistic studies, that is secondary genres. As such, the study hopefully provides reliable insights into the very specific domain of legal interlingual communication. The operationalization of the process is systemic, and it takes account of well-established categories, like discrete units and CQL data extraction, which were followed by thorough statistical processing. The author is aware of the limitations of the study in that (i) the sample for parallel, qualitative analysis is small (ii) the discrete units are not always held as reliable markers of stylistics, although the rationale for this methodology was accounted for, and (iii) statistical significance tests need to be applied for contrasting all of the parameters.

The project can certainly be extended by conducting a more in-depth parallel search. This would develop the qualitative aspect of the analysis and enable the results to be compared to a number of concepts on the ground of translation studies, such as comprehensibility and readability studies, quality assessments and evaluation of translation errors. This could also constitute the ground for developing systemic models for translation performance, taking account of the optimal, usage-based readability factor.

Bibliography

Bázlik M., Ambrus P., *A Grammar of Legal English*, Bratislava 2008.
Biel Ł., *Lost in the Eurofog: the Textual Fit of Translated Law*, Frankfurt am Main 2014.

Biel Ł., *Phraseological profiles of legislative genres: Complex prepositions as a special case of legal phrasemes in EU law and national law*, "Fachsprache" 2015, no. 3–4, pp. 139–160.

Dodsworth R., *Speech communities, social networks, and communities of practice*, [in:] *Research Methods in Sociolinguistics. A Practical Guide*, J. Holmes, K. Hazen (eds.), Hoboken 2014.

Gotti M., *Investigating Specialised Discourse*, Bern 2005.

Goźdź-Roszkowski S., *Patterns in Linguistic Variation in American Legal English*, Frankfurt am Main 2011.

Gumul E., *Explicitation in Simultaneous Interpreting*, Katowice 2017.

Hiltunen R., *The grammar and structure of legal texts*, [in:] *Language and Law*, T. Peter, S. Lawrence (eds.), Oxford 2012, pp. 39–52.

House J., *English as a lingua franca and translation*, "The Interpreter and Translator Trainer" 2013, no. 7(2), pp. 279–298.

Kjær A.L., *Phrasemes in legal texts*, [in:] *Ein Internationales Handbuch der Zeitgenössischen Forschung/Phraseology. An International Handbook of Contemporary Research*, Berlin – New York 2007, pp. 506–515.

Lapshinova-Koltunski E., *Variation in translation: Evidence Corpora*, [in:] *New directions in corpus-based translation studies*, F. Fantinuli, C. Zanettin (eds.), Berlin 2015, pp. 93–114.

Legallois D., Charnois T., Larjavaara M., *The balance between quantitative and qualitative literary stylistics: how the method of "motifs" can help*, [in:] *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, D. Legallois, T. Charnois, M. Larjavaara (eds.), Berlin 2018, pp. 165–193.

Lewandowska-Tomaszczyk B. (ed.), *Podstawy językoznawstwa korpusowego*, Łodź 2005.

Longerée D., Mellet S., *Towards a topological grammar of genres and styles: a way to combine paradigmatic quantitative analysis with a syntagmatic approach*, [in:] *The Grammar of Genres and Styles: From Discrete to Non-Discrete Units*, D. Legallois, T. Charnois, M. Larjavaara (eds.), Berlin 2018, pp. 140–163.

Melchers G., Shaw P., Sundkvist P., *World Englishes*, New York 2019.

Olohan M., *How frequent are contractions? A study of contracted forms in the Translation English Corpus*, "Target" 2003, no. 15, pp. 59–89.

Peruzzo K., *Legal system: an additional variable in the analysis of short-term diachronic evolution of legal terminology*, "International Journal of Legal Discourse" 2017, no. 2(2), pp. 291–313.

Rabadán R., *Translating the "predictive" and "hypothetical" meanings English/Spanish*, "Meta" 2007, no. 52(3), pp. 484–502.

Salkie R., *Legal phraseology in contrast. The fact that and its German counterparts*, [in:] *Phraseology in Legal and Institutional Settings. A Corpus- based Interdisciplinary Perspective*, S. Goźdź Roszkowski, G. Pontrandolfo (eds.), London – New York 2018, pp. 126–143.

Szymor N., *Behavioural Profiling in Translation Studies*, "Trans-Kom" 2015, no. 8(2).

Trosborg A., *Rhetorical Strategies in Legal Language*, Tuebingen 1997.

Tyler A., Evans V., *The Semantics of English Prepositions. Spatial Scenes, Embodied Meaning and Cognition*, Cambridge 2003.

Vanderbauwhede G., Desmet P., Lauwers P., *The shifting of the demonstrative determiner in French and Dutch in parallel corpora: From translation mechanisms to structural differences*, "Meta" 2011, no. 56(2), pp. 443–464.

Vigier F.J., Mar Sánchez Ramos M. del, *Using parallel corpora to study the translation of legal system-bound terms: The case of names of English and Spanish courts*, [in:] *Computational and Corpus-based Phraseology. Second International Conference, Europhras 2017 London UK, November 13-14, 2017 Proceedings*, Cham 2017.

Więcławska E., *On various degrees of stability of multi-word combinations in the language of Polish law*, "International English Studies Journal, Studia Anglica Resoviensia" 2013, no. 10, pp. 73–86.

Więcławska E., *Capitalising on translation market data in the field of commercial law*, "Comparative Legilinguistics" 2018, no. 33, pp. 85–115.

Więcławska E., *English/Polish contrasts in legal language from the usage-based perspective*, [in:] *Redefining and Refocusing Translation and Interpreting Studies Selected Papers form the 3rd International Conference on Translation and Interpreting Studies TRANSLATA III (Insbruck 2017)*, L. Lanthaler, R. Lukenda (eds.), Berlin 2020, pp. 99–104.

Więcławska E., *Extrapolating the adjective distribution in the Polish legal language*, [in:] *Specialist Languages in Use and Translation*, Rzeszów 2014, pp. 71–82.

Więcławska E., *Field-specific conventions in the translation of company law documentation for court proceedings*, "Studies in Logic, Grammar and Rhetoric" 2019, no. 58(71), pp. 221–243.

Więcławska E., *Language of law: Court translation in the didactic perspective*, [in:] *Specialist Communication. Student Training and Market Demands*, Rzeszów 2018, pp. 37–60.

Więcławska E., *Sociolinguistic and Grammatical Aspects of English Company Registration Discourse*, "Humanities and Social Sciences" 2019, vol. XXIV, no. 26(4).

Więcławska E., *Towards the establishment of genre distinctions of company legal registration proceedings*, "Komunikacja Specjalistyczna" 2018, no. 16, pp. 145–167.

Wiesmann E., *Rechtsübersetzung und Hilfsmittel zur Translation*, Tübingen 2004.

Williams C., *Tradition and Change in Legal English. Vernal Constructions in Prescriptive Texts*, Bern 2005.

Wolfer S., *Comprehension and comprehensibility*, [in:] *Translation and Comprehensibility*, K. Maksymski, S. Gutermuth, S. Hansen-Schirra (eds.), Berlin 2015, pp. 33–53.