

LEXICAL BUNDLES IN THE CORPUS OF SLOVAK JUDICIAL DECISIONS

MIROSLAV ZUMRÍK

Slovak National Corpus, L. Štúr Institute of Linguistics, Slovak Academy of Sciences,
Bratislava, Slovakia

ZUMRÍK, Miroslav: Lexical bundles in the corpus of Slovak judicial decisions. *Journal of Linguistics*, 2021, Vol. 72, No 2, pp. 705 – 718.

Abstract: The paper follows the tradition of research in legal linguistics and into formulaic language, specifically into lexical bundles. The aim of the paper is to describe lexical bundles in samples from the corpus of Slovak judicial decisions OD-JUSTICE by means of quantitative characteristics of the identified bundles and by their comparison with bundles found in two other specialized corpora: the corpus of Slovak legal regulations and the corpus of annual reports by Slovak public institutions. For the identification of bundles, the concept of the h-point was used. Identified bundles are described with respect to their maximal, minimal, average, median and mode values, distributions and ratios. The aim of the paper is to outline an interpretation of these bundle characteristics with regard to communicative function(s) of compared document genres.

Keywords: lexical bundles, formulaic language, judicial decisions, specialized discourse, legal linguistics, pattern-driven research

1 INTRODUCTION

That a natural connection between language and law already exists can be deduced from the fact that language, according to J. Prusák, serves as a presupposition for the existence of the legal norms in which they are written [1, p. 295]. This relationship brings about another close connection, namely that which exists between law and linguistics. Their interaction, termed legal linguistics, can be seen as a “mutual arena for cooperation” [2], where one’s interest in law may justifiably imply an interest in the nuances and peculiarities of legal language [3, p. 1]. Following the tradition of Slovak legal linguistics pioneered by Š. Luby or R. Kuchár, an interest in the linguistic aspects of Slovak legal language nowadays covers a range of areas, from the stylistic features of (mostly) legal norms ([3], [4]) to anaphoric and other tools for creating concise legal texts [5].

One phenomenological observation of legal language might be that legal texts tend to contain recurrent word sequences. This linguistic phenomenon has been studied as so called formulaic language, or formulaicity, which is often understood as characteristic of legal discourse [6, p. 7]. Such co-occurrence of language units

constituting multi-word expressions is, however, a feature present in the language in general – a systemic feature or an “axiom” [7, p. 10]. Formulaicity is further conceptualized as a pole on a continuous scale of qualities, as opposed to the pole of free creativity [8]. Given its scale-like nature, formulaicity can be studied and expressed quantitatively, similar to qualities of terms [9] or translated texts [10]. In other words, the formulaicity can be measured and therefore can be searched for possible “formulas” for formulaic structures [11], which can cover up to 40 types, from amalgams to unanalysed portions of speech [12, p. 3].

2 THE STUDY OF LEXICAL BUNDLES AND THE RESEARCH AIM

One of these formulaic structures in linguistics has been studied as so-called *lexical bundles* [13], such as ‘*on the other hand*’, ‘*as can be seen*’ or ‘*it is recommended that*’. Researchers focus predominantly on more or less specialized discourses: academic discourse [14], medical leaflets [15] and legal texts, the latter from perspectives such as genre [16], linguistic structure [6], translation strategies [17] or legal semantics [18].

Lexical bundles have traditionally, by D. Biber, been defined on the basis of frequency as “sequences of word forms that commonly go together in natural discourse” and that “show statistical tendency to co-occur” [13, pp. 989–990]. Originally, three criteria have been used to define what classifies as a lexical bundle: these are, the minimal orthographic length of a sequence (3–8 tokens), its minimal normalized frequency (this cut-off point varies in individual approaches between 10, 20, 40 or more) and the “dispersion range” of a bundle throughout the individual texts in a given sample – this value could be, for example, the minimal appearance of a bundle in 5 different documents [13] or in 10 percent of sample texts [19].

Lexical bundles and other formulaic sequences are considered to be the building blocks of a given discourse ([20], [6]) which improve processing efficiency in communication [21]. One’s command of such multi-word sequences is also said to have a sociological value for signalling an individual’s belonging to a community [12] or a pedagogical value in teaching and mastering of specialized discourses [22].

Since the introduction of lexical bundles as a term, a methodological discussion has also emerged regarding a “fine tuning” of the process by which bundles are identified [23]. This involves searching for a method which would not only be based on arbitrary set cut-off points, but one that would uncover bundles more or less typical of a given discourse. Among such methods, Ł. Grabowski proposes the use of multiple sampling techniques, and, even where such bundles are identified as the most frequent and often structurally incomplete word sequences, he recommends that efforts to be directed at ways to identify preferably “structurally complete” bundles which can more easily be ascribed discourse function(s) [23, pp. 63–68].

Bearing the need for such methodological “fine-tuning” in mind, I would like to apply the method adopted in Cvrček et al. [24], where the formulaicity of texts is

expressed as the number of those 5-grams that have a frequency higher than the so-called h-point, introduced by I. I. Popescu [24, pp. 63 and 75]. The rationale for using the concept of h-point is that this point divides words into two areas: a branch of more frequent synsemantic words, and a branch of less frequent autosemantic words. If an autosemantic word, according to its frequency, appears in the synsemantic area, such a word can be perceived as “some kind of anomaly in comparison to ‘neutral’ texts” [25, pp. 217–218], that is, in some way specific to the text from which it originates. The h-point can be defined as the point where the rank of a given word equals its frequency. Where there is no such word, the h-point is calculated as follows, where $r(\text{ank})_1 > r(\text{ank})_2$ and $f(\text{frequency})_1 > f(\text{frequency})_2$:

$$(f_1 r_2 - f_2 r_1) / ((r_2 - r_1) + f_1 - f_2)$$

By using the h-point calculations and quantitative characteristics, such as maximum, minimum, average, median or mode values, distribution of bundles throughout documents and corpora, as well as qualitative characteristics of the bundles (their communicative function), the aim of the paper is to describe lexical bundles in the corpus of Slovak decisions by means of comparison with bundles found in the corpora of Slovak legal acts and annual reports by Slovak public institutions. In other words, the paper is an attempt to find out whether, and to what degree, these quantitative characteristics correspond with the nature of both documents and institutional procedures from which the documents result.

3 DATA

The material basis for the research consisted of three specialized corpora:

1. The rather large corpus of Slovak judicial decisions od-justice-1.0 (further referred to as OD).	4,149,442,677 tokens	2,622,795 documents
2. The corpus of Slovak legal regulations (Korpus slovenských právnych predpisov k 2020-12-01, v1.4 , further referred to as A(cts)).	43,750,050 tokens	20,186 documents
3. The corpus of annual reports by Slovak public institutions gov-vs-1.0 (further referred to as AR). Here, it should be noted that the data had been de-duplicated at the level of paragraphs, which needs to be taken into consideration.	17,864,463 tokens	1016 documents

Tab. 1.

Corpora 1 and 3 are available via the Department of Slovak National Corpus at <https://bonito.korpus.sk>, corpus 2 via the webpage of the Ľ. Štúr Institute of Linguistics at <https://alica.juls.savba.sk>.

Because the magnitude of corpora caused technical problems regarding the search for relevant bundles within its range, it was decided to work with two samples from each corpus, one smaller, at approx. 2 million tokens and one larger at approx. 10 million. The aim was to work with large portions of data of similar dimensions from all corpora.

In the corpus of judicial decisions, the verdicts can be grouped and searched in accordance with the date on which they were announced. This means that one can choose a subcorpus of decisions which only have this date in common. For analysis of decisions, I therefore picked 1000 documents from 10 July 2018 and 4968 documents from 29 June 2016.

Similarly, documents in the corpus of legal regulations are organized according to the year of their promulgation, but also according to subgenre (act, decree, resolution, regulation etc.). I decided to only work with acts, where the smaller sample were 472 documents from 2018 and 2019, and the larger sample of 2561 documents from the years 2007–2017 and 2020.

The annual reports are, understandably, sorted according to the years which correspond with the activities of the institutions they summarize. Here, the smaller sample consisted of 110 reports from 2012, while the larger one consists of 536 reports from 2007–2011, 2013 and 2014.

4 METHOD

From the 6 samples (two for each corpus, one smaller and referred to as OD/A/AR, one larger, referred to as ODext/Aext/ARext), it was necessary to identify lexical bundles above the respective h-point in each of the samples. Lexical bundles were identified using the CQL search in NoSketch Engine, as 5 consecutive tokens within a sentence (the latter condition so that the 5-token window would not take in sequences of words from neighbouring sentences). In order to further avoid counting punctuation, symbols or numbers as words, the tokens had to be attributed a morphological annotation in the range S (noun) – Y (conditional morpheme), thus excluding W (abbreviation, symbol) – 0 (number).

The identified 5-grams were ranked according to their absolute frequency. The next step was to find the h-point of a given sample: with ranks of individual bundles already identified, a search was made for the bundle that would have the same rank and the (normalized) frequency (*ipm*). In this way, the h-point for a given sample was either found directly (in 5 cases), or calculated using the formula mentioned in Section 2 (in 1 case).

The bundles above the h-point were sorted manually into groups, consisting of either

1. at least partially overlapping 5-word bundles, which constitute parts of longer bundles, as in *len do uplynutia lehoty na* ‘only until expiration of (...) period’¹ – *do uplynutia lehoty na podanie* ‘until expiration of (appellate period)’, or

2. bundles with similar, albeit not identical wording, as in *Poznámka pod čiarou k odkazu* ‘Footnote with respect to reference’ – *Poznámky pod čiarou k odkazu* ‘Footnotes with respect to reference’.

The issue of dealing with overlapping bundles is addressed in Grabowski [23, p. 63–67] where several possible approaches are mentioned. Here, the groups were identified manually by looking at the frequent right and left concordances of a bundle. The values of h-points and list of bundles above the h-point in each sample were then arranged in 6 tables for further processing, that is, counting and sorting overlapping and similar bundles within groups. Apart from bundles in groups (consisting of 2- to 12 bundles), those not contained within longer ones (thus representing genuine 5-grams) were also counted.

The 6 sample tables further contained values of normalized frequencies for each bundle, their dispersion throughout the documents, the ratio between the dispersion and number of sample documents in total. For these characteristics, the maximum, minimum, average, median and mode values were calculated. The values are then arranged in tables that serve as the starting point for the findings in Section 5.

5 FINDINGS

5.1 Maximal, minimal and average values of tokens per documents ratio in the samples (*tok/doc*)

	OD	ODext	A	Aext	AR	ARext
tokens	1,846,380	10,092,069	2,233,571	10,952,481	1,889,037	10,185,459
documents	1,000	4,968	472	2,561	110	536
max tok/doc	13,400	27,025	121,796	186,700	144,857	127,152
min tok/doc	239	249	109	97	1,932	1,652
avr tok/doc	1,846	2,031	4,732	4,276	17,173	19,003

Tab. 2.

The values vary in all three corpora, while the span is larger in the A sample. This makes the minimum values in A more similar to that of OD, while the maximum in A is more similar to AR. Even though the total numbers of tokens in both samples from all three corpora, as well as the ratio between token count in smaller and larger sample remain approximately the same (1:5), the number of documents in corpora differs, which makes average document length unequal. Comparison of smaller and larger samples shows that their average document length is approximately the same.

¹ I would like to thank Juraj Kotrusz for translation of some of the legal lexical bundles.

5.2 Number of lexical bundles above the h-point (*LBs > h-point*)

	OD	ODext	A	Aext	AR	ARext
LBs > h-point	106	85	55	42	33	26

Tab. 3.

This value remains approximately the same within each of the corpora (106 and 85 in OD, 55 and 42 in A, 33 and 26 in AR) while the ratio between those corpora can be described as 1 – ½ – 1/3. This would seem to indicate a higher degree of formulaicity found in the OD corpus, however, it needs to be considered that there are half as much documents in OD. The number of bundles in the smaller A sample is approximately one third the number in the smaller OD sample, but the bundles from A appear in ten times more documents.

5.3 Number of groups of (partially) overlapping or similar lexical bundles (*LB groups*), identified manually

	OD	ODext	A	Aext	AR	ARext
LB groups	29	29	14	16	16	13

Tab. 4.

Similarly, to point 2, the distribution of groups within corpora remains more or less the same, with 29 groups in OD and 13, 14 or 16 groups in both A and AR. Again, this could indicate a bigger “diversity” of formulaic sequence types in the OD corpus, but the appearance might also be related to different average document lengths in the three corpora, as mentioned in section 5.1.

5.4 Distribution of bundle groups (*dist bnd grps*)

Described as x(y), where y is the number of 5-word bundles within the group and x is the number of groups. Examples of frequent longer sequences comprised of partially overlapping five-word bundles in each of the six samples can be seen below the respective numbers of bundle groups, with the first (most frequent) five-word bundle marked in bold. The number of groups in each sample needs to be completed with the number of wording variations in Tab. 5.

	OD	ODext	A	Aext	AR	ARext
dist bnd grps	2(8)	1(8)	x	x	1(8)	1(8)
	2(7)	x	1(10)	x	x	x
	1(6)	2(6)	x	x	x	x

	OD	ODext	A	Aext	AR	ARext
	2(5)	4(5)	1(5)	1(5)	x	x
	6(4)	3(4)	2(4)	x	x	x
	3(3)	4(3)	x	x	3(3)	1(3)
	3(2)	6(2)	3(2)	3(2)	4(2)	4(2)
	7(1)	9(1)	4(1)	9(1)	8(1)	7(1)

Tab. 5.

The almost 30 groups in both OD samples are distributed relatively homogeneously with 1 to even 6 occurrences of bundles consisting of eight, seven, six etc. overlapping 5-word bundles. In samples from the AR corpus, we can notice appearance of groups consisting of 8 bundles. But apart from that, groups of 5 to 2 bundles appear only 1 to 4 times, leading to the conclusion that the distribution in the middle group size range is more heterogeneous here. Bundles consisting of just 5 words, however, appear with similar frequency in most samples, with the exception of the smaller A sample. This observation would suggest that bundles in OD corpus are both more diverse and that they appear with equal regularity, while there is a lower bundle diversity in the A and AR corpora, where the bundles, in turn, constitute more substantial groups (longer sequences).

5.5 Number of similar bundles within a group (*incl word var*)

Described as $x(y/w-z\dots)$, where y represents the number of bundles within the group, $w, z\dots$ are the numbers of bundles containing word variations in the group and x is the number of occurrences for y . Examples of wording variation found in all six samples can be seen below the numerical variation schemes.

	OD	ODext	A	Aext	AR	ARext
incl word var	1(8/3-4-1)	x	1(12/1-7-1-3)	1(12/1-4-1-3-1-2)	x	x
	1(4/2-2)		1(7/2-2-1-2)	1(8/2-2-3-1)		
	1(2/1-1)		1(2/1-1)	1(2/1-1)		

Tab. 5. cont.

This is a subsection of groups identified in 5.4, and here one observes a similarity between the number of wording variations found in both OD and AR corpora. Both A samples contained several groups with more variations and wording similarities. This might be related to the presence of less diverse, albeit longer, bundle sequences especially in the A/Aext corpus, as described in point 5.4.

5.6 Maximal, minimal, average, median and mode values for normalized frequency of individual bundles (*ipm*)

	OD	ODext	A	Aext	AR	ARext
max <i>ipm</i>	286.5	274.9	843.0	670.1	434.6	290.0
min <i>ipm</i>	106.7	86	57.8	42.3	33.8	26.8
avr <i>ipm</i>	115.7	166.7	179.1	166.8	100.5	81.2
med <i>ipm</i>	139.7	160.8	100.7	116.2	45.5	40.6
mod <i>ipm</i>	132.7	160.8	127.1	129.9	33.8	35.2

Tab. 6.

Maximal and minimal normalized frequencies remain relatively similar in smaller and larger samples within all three corpora, while the maximum value in OD samples is approximately four times lower than that in A samples, and almost the same/one third lower than in the AR samples. On the contrary, minimal values in OD samples are higher than in both A (two-) and AR corpora (three times). This could be perceived as another trace of distributional structures within bundle groups as described in point 5.4, with more homogeneous distribution of various and more numerous bundle groups in OD samples. However, average *ipm* values in all three corpora are relatively similar, while there is a bigger similarity between median and mode values in OD and A corpora than in the AR corpus.

5.7 Maximal, minimal, average, median and mode values of bundle dispersion per sample ratio (*dis/s*)

Calculated as dispersion value divided by the number of all documents within a sample. Dispersion (*dis*) is calculated as the number of first-time appearances of a given bundle in sample documents, that is, as the number of documents in the sample where the bundle appears at least once.

	OD	ODext	A	Aext	AR	ARext
max <i>dis/s</i>	50,3	53,8	60,1	55,7	83,6	65,9
min <i>dis/s</i>	6,7	8,5	2,8	0,04	0,9	0,9
avr <i>dis/s</i>	24,7	29,3	21,7	26,2	28,3	31,2
med <i>dis/s</i>	25,1	32,3	16,1	21,2	20	32,2
mod <i>dis/s</i>	24,5	32,4	4,9	55,6	1,8	0,9

Tab. 7.

In this respect, individual bundles appear to the maximum value in around 50 percent of OD sample documents, whereas the maximum values for normalized maximal dispersion are slightly higher (50–60 percent) in A samples, and even higher (more than 60/80 percent) in AR samples. Minimal values seem to behave in an inverted order, with the highest minimal value in OD and lowest in AR samples.

Both average and median values for the normalized dispersion, however, are relatively similar in all six samples. The mode values decrease in smaller OD, A and AR samples, respectively, while this value is relatively higher in larger OD and A samples, but below 1 in the larger AR sample.

5.8 Communicative functions of the bundles

Apart from distributional and proportional characteristics summarized in the table above, the 6 overview sample tables enabled to compare bundles in three corpora according to communicative (discourse) functions of bundles. The question would be, how to compare communicative functions of bundles in different genres. Here, it is possible to use a common classification scheme of these functions, as the one originally used by D. Biber et al. (stance bundles, discourse organizers, referential bundles) [20]. It is also possible to group the present bundles in each genre according to the more or less schematic textual structure of administrative texts, such as decisions, acts, reports.

5.8.1 Communicative functions of the bundles found in judicial decisions

Communicative functions of bundles found in judicial decisions can be described by placing the individual bundle into four categories, corresponding to basic parts of a decision, where these bundles normally appear. These parts are

- a) heading part which identifies the case, the involved parties and other circumstances;
- b) enunciation/verdict part which pronounces the verdict, often in several counts;
- c) reasoning part which presents the reasons that have led the court to reach its verdict;
- d) instruction part which advises the party suffering as a result of the decision on possible remedies.

Following this textual structure, the lexical bundle groups found in both samples of OD corpus fall in line with distributional characteristics presented in point 5.4 mostly in instruction and reasoning parts.

Communicative function/Type	LB groups in OD sample	Examples	ipm
instruction part	15	1. <i>len do uplynutia lehoty na (podanie odvolania)</i> 'only until expiration of (appellate) period'	286
		2. <i>z akých dôvodov sa rozhodnutie (považuje za nesprávne)</i> '(what are the) grounds for (considering) the decision (as incorrect)'	272
		3. <i>v akom rozsahu sa (rozhodnutie) napáda</i> '(what is) the extent of appeal'	269

Communicative function/Type	LB groups in OD sample	Examples	ipm
reasoning part	10	1. (o) súdnych exekútoroch a exekučnej činnosti 'on judicial enforcers and on enforcement proceeding'	241
		2. a náhrada za stratu času 'and compensation of lost time'	206
		3. v konkurze alebo splátkovým kalendárom 'bankrupt or by repayment plan'	173
verdict part	3	1. nárok na náhradu trov konania 'entitlement to covering of costs of enforcement proceeding'	186
		2. odo dňa doručenia platobného rozkazu 'after the service of the charging order in writing'	168
		3. (ktoré) môžu byť uspokojené iba v '(which) can only be met in'	122
heading part	0		
other	1	MENE SLOVENSKEJ REPUBLIKY Okresný súd 'BEHALF OF THE SLOVAK REPUBLIC District court'	121
Communicative function/Type	LB groups in ODext sample	Examples	ipm
instruction part	19	1. a čoho sa odvolateľ domáha 'what is pursued by the appellant'	266
		2. alebo postup súdu považuje za (nesprávny) 'or considers court's procedural measures (to be unlawful)'	259
		3. v čom sa toto rozhodnutie '(what are the grounds for considering) this decision'	253
reasoning part	7	1. (o) udelenie poverenia na vykonanie exekúcie '(for) granting of authorization for enforcement'	274
		2. právo na náhradu trov konania 'entitlement to covering of costs of enforcement proceeding'	205
		3. (o) súdnych exekútoroch a exekučnej činnosti 'on judicial enforcers and on enforcement proceeding'	142
verdict part	2	1. nemá právo na náhradu trov 'is not entitled to covering of costs of enforcement proceeding'	134
		2. (s) úrokom z omeškania vo výške '(with) late charges of'	107
heading part	0		
other	1	MENE SLOVENSKEJ REPUBLIKY Okresný súd 'BEHALF OF THE SLOVAK REPUBLIC District court'	97

Tab. 8.

5.8.2 Communicative functions of the bundles found in acts

The 5-word bundle groups found in acts are mostly parts of formulas, above all the amendment and supplementing formula, consisting of twelve 5-word bundles (including variations); promulgation formula, through which an act comes into being, so to speak; or formulas referring to the body of legal text itself, signalling a footnote or a change to the text.

Communicative function/Type	LB groups in A/Aext sample	Examples	ipm
amendment and supplementing formula	3/3	1. <i>a o zmene a doplnení (niektorých zákonov)</i> 'and on amendment and supplementing of (several statutes)'	843/670
		2. <i>sa mení a dopĺňa takto</i> 'is amended and supplemented in the following way'	316/241
		3. <i>ktorým sa mení a dopĺňa (zákon)</i> 'by which (the statute) is amended and supplemented'	106/123
text reference formula	2/3	1. <i>Poznámka pod čiarou k odkazu</i> 'Footnote with respect to reference'	650/528
		2. <i>Poznámky pod čiarou k odkazu</i> 'Footnotes with respect to reference'	198/174
		3. <i>sa na konci pripájajú tieto (slová)</i> 'following (words) are supplemented to the end'	172/149
promulgation formula	1/1	<i>(Národná rada Slovenskej republiky) sa uzniesla na tomto zákone</i> '(The National Council of the Slovak Republic) has enacted the following statute'	127/130
other	8/9	1. <i>nálezu Ústavného súdu Slovenskej republiky</i> '(of the) finding of the Constitutional Court of the Slovak Republic'	145/114
		2. <i>(sociálnoprávnej) ochrany detí a sociálnej kurately</i> 'of the child welfare services'	89 (A)
		3. <i>ak tento zákon neustanovuje inak</i> 'unless this Act stipulates otherwise'	76 (Aext)

Tab. 9.

5.8.3 Communicative functions of the bundles found in annual reports

The most common formula (amendment and supplementing f.) found in acts is also that which is most commonly found in annual reports (here, it consist of 8 shorter bundles), while, apart from that, bundles in the annual reports consist to some extent of proper names referred to in the report texts.

Communicative function/Type	LB groups in AR/ARext sample	Examples	ipm
amendment and supplementing formula	2/2	1. <i>a o zmene a doplnení (niektorých zákonov)</i> 'and on amendment and supplementing of (several statutes)'	434/290
		2. <i>ktorým sa mení a dopĺňa (zákon)</i> 'by which (statute) is amended and supplemented'	93/69
proper name	3/2	1. <i>Ministerstva pôdohospodárstva a rozvoja vidieka (Slovenskej republiky)</i> '(of the) Ministry of Agriculture and Rural Development (of the Slovak Republic)'	50 (AR)
		2. <i>Štátna vedecká knižnica v Prešove</i> 'State Science Library in Prešov'	40/27
		3. <i>Ministerstva životného prostredia Slovenskej republiky</i> '(of the) Ministry of Environment of the Slovak Republic'	35/26

Communicative function/Type	LB groups in AR/ARext sample	Examples	ipm
other	11/9	1. <i>(pri) výkone práce vo verejnom záujme</i> 'during the public interest service' 2. <i>škôl a školských zariadení v</i> '(of) schools and school facilities in' 3. <i>štátnej správy starostlivosti o životné prostredie</i> '(of) national environmental administration'	66/63 59/46 43/31

Tab. 10.

6 CONCLUSION

The aim in presenting the quantitative characteristics was not to prove that judicial decisions are simply more or less formulaic (that is, schematic or prefabricated) than acts or annual reports. Every text, register, style or genre has its own means by which it can be considered to accomplish its communicative function, even when these methods can become subject to dispute, because they can be perceived as not sufficiently effective or stylistically balanced, as F. Štícha points out with reference to the style of judicial decisions [26, pp. 71–72]. There is no “linear”, deterministic connection between the use of certain lexis in a given text and its communicative function. Nor can we afford to neglect the existence of “style-mixes” and “transitional areas” ([27], cited in [4, p. 208]). This is especially true in the case of judicial decisions, which often contain the explicit language given in testimonies, as well as specialized legal terms and analytic multi-word expressions. This opens for the possibility that the feature called formulaicity, as well as complex linguistic phenomena in general, might be productively studied by applying methods of multidimensional analysis, such as this has been showcased in [24].

Nevertheless, the lexical bundles found in samples of Slovak judicial decisions are word sequences that constitute the textual result of an institutionally regulated social interaction between these institutions (courts) and involved parties, be it physical and/or juridical personae. The decisions are thus directed both at the realms of “normativity” [28, pp. 84–113] and factuality, as the decisions represent a multifaceted, possibly complex [29, p. 216] process of law application (the realm of regulations and norms) regarding the factual case (involving individuals or legal entities). As shown in Section 5, point 9, a noticeable quantity of bundles found in Slovak judicial decisions informs involved parties when deciding how to react to the implications of a court ruling (e. g. on how to appeal) and upon what rationale has the court based its verdict. The intersection of individuals and individual cases (factuality) and regulations (normativity) might, then, require a set of detailed regulations (mostly sections 363, 364 and 365 of the Slovak Civil Contentious Procedure Code), that constitute the reasoning and instructional components of its

decisions. This, in turn, may have resulted in the relative high diversity and relative low variation of bundles in Slovak judicial decisions, as compared to acts.

The acts, on the other hand, address not so much particular individuals (even though they are produced in institutional settings of legislative bodies) as much as decisions arrived at in courts. The acts inform the broader public regarding how social interaction in various domains should play out, while also outlining the implications (sanctions) for non-compliance with these regulations. This means that acts express both norms and a model-like representation of reality, while operating more in the realm of norms and ideals, in that they both incorporate new regulations into the body of legislation and change existing regulations. Lexical bundles found in the act samples bear witness of these legislative procedures, as they consist mainly of traditional formulae used when referring to the actual and/or amended wording of acts, or formulas for the promulgation of new regulations. The aforementioned reference formulae are also found as bundles in Slovak annual reports, while there are also bundles denoting proper names of institutions or document titles. These bundles constitute (perhaps because they are more bound to specific report subjects) shorter bundle groups. Some future research into patterns emerging via formulaic sequences in texts produced in institutional settings might constitute a promising field of study.

ACKNOWLEDGEMENTS

The paper has been written within the Slovak National Corpus project supported by the Slovak Academy of Sciences, Ministry of Education, Science, Research and Sport of the Slovak Republic, Ministry of Culture of the Slovak Republic and the Ľudovít Štúr Institute of Linguistics, Slovak Academy of Sciences.

References

- [1] Prusák, J. (2001). *Teória práva*. Bratislava: Vydavateľské oddelenie Právnickej fakulty UK 2001, 331 p.
- [2] Engberg, J. (2013). Legal linguistics as a mutual arena for cooperation. *Recent developments in the field of applied linguistics and law*. AILA Review 26, pages 24–41.
- [3] Holländer, P. (2014). *Interpretácia práva*. Justičná akadémia. Available at: https://www.ja-sr.sk/files/Interpretacia_prava.pdf.
- [4] Imrichová, M. (2014). Jazyk právnych textov so zreteľom na špecifiká právnej lexiky. In J. Kesselová, M. Imrichová and M. Ološtiak (eds.), *Registre jazyka a jazykovedy (I)*. Prešov. Filozofická fakulta Prešovskej univerzity v Prešove, pages 207–212.
- [5] Gahér, F., Števec, M., and Braxatoris, M. (2019). Nástroje a pravidlá produkcie a interpretácie koncízneho textu (s osobitným zreteľom na normatívu). *Jazykovedný časopis*, 70(1), pages 73–94.
- [6] Berūkštienė, D. (2017). A corpus-driven analysis of structural types of lexical bundles in court judgments in English and their translation into Lithuanian. *Kalbotyra* 70, pages 7–31.
- [7] Kačala, J. (2010). *Zložené útvary v jazyku*. Martin: Vydavateľstvo Matice slovenskej. 134 p.
- [8] I. MacKenzie and M. A. Kayman (eds.). (2018). *Formulaicity and Creativity in Language and Literature*. Routledge. 126 p.

- [9] Kovářiková, D. (2017). *Kvantitativní charakteristiky termínů*. Praha: Nakladatelství Lidové noviny. 136 p.
- [10] Chlumská, L. (2017). *Překladová čeština a její charakteristiky*. Praha: Nakladatelství Lidové noviny. 150 p.
- [11] Grabowski, Ł., and Forsyth, R. S. (2015). Is there a formula for formulaic language? *Poznan Studies in Contemporary Linguistics* 51(4), pages 511–549.
- [12] Wray, A., and Perkins, M. (2000). The functions of formulaic language: An integrated model. *Language & Communication* 20(1), pages 1–28.
- [13] Biber, D., Johansson S., Leech G., Conrad S., and Finegan E. (1999). *The Longman Grammar of Spoken and Written English*. Harlow (Essex): Longman, 1204 p.
- [14] Biber, D., and Barbieri, F. (2007). Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26, pages 263–286.
- [15] Grabowski, Ł. (2014). On lexical bundles in Polish patient information leaflets: a corpus-driven analysis. *Studies in Polish Linguistics* 9(1), pages 21–43.
- [16] Breeze, R. (2013). Lexical bundles across four legal genres. *International Journal of Corpus Linguistics* 18(2), pages 229–253.
- [17] Biel, Ł. (2017). Lexical bundles in EU law: the impact of translation process on the patterning of legal language. In *Phraseology in legal and institutional settings. A corpus-based interdisciplinary perspective (e-book)*. S. Goźdz-Roszkowski and G. Pontrandolfo (eds.). London, New York: Routledge, pages 10–26.
- [18] Goźdz-Roszkowski, S. (2018). Between corpus-based and corpus-driven approaches to textual recurrence: Exploring semantic sequences in judicial discourse. In J. Kopaczyk and J. Tyrkkö (eds.), *Applications of Pattern-driven Methods in Corpus Linguistics*, pages 131–158.
- [19] Hyland, K. (2008). As can be seen: lexical bundles and disciplinary variation. *English for Specific Purposes* 27, pages 4–21.
- [20] Biber, D., Conrad, S., and Cortes, V. (2004). If you look at...: lexical bundles in university teaching and textbooks. *Applied Linguistics* 25 (3), pages 371–405.
- [21] Wiechmann, D., and Kerz, E. (2016). Formulaicity as a determinant of processing efficiency: investigating clause ordering in complex sentences. *English Language and Linguistics* 20(3), pages 421–437.
- [22] Tománková, V. (2016). Lexical bundles in legal texts corpora – selection, classification and pedagogical implications. *Discourse and interaction* 9(2), pages 75–94.
- [23] Grabowski, Ł. (2018). Fine-tuning lexical bundles: A methodological reflection in the context of describing drug-drug interactions. In J. Kopaczyk and J. Tyrkkö (eds.), *Applications of Pattern-driven Methods in Corpus Linguistics*, pages 15–56.
- [24] Cvrček, V., Laubeová, Z., Lukeš, D., Poukarová, P., Řehořková, A., and Zasina, A. J. (2020). *Registry v češtině*. Praha: Nakladatelství Lidové noviny. 234 p.
- [25] Čech, R., Garabík, R., and Altmann, G. (2015). Testing the Thematic Concentration of Text, *Journal of Quantitative Linguistics*, 22(3), pages 215–232.
- [26] Štícha, F. (1985). O jazyce soudních rozhodnutí. *Naše řeč*, 68(2), pages 68–77.
- [27] Slančová, D. (1999). *Reč autority a lásky: reč učitelky materskej školy orientovaná na dieťa*. Opis registra. Prešov: Filozofická fakulta Prešovskej univerzity. 224 p.
- [28] Procházka, R., and Káčer, M. (2019). *Teória práva*. 2. vydanie. Bratislava: C. H. Beck. 304 p.
- [29] Gábriš, T. (2018). Sudcovské rozhodovanie z pohľadu súčasného právneho realizmu. In T. Gábriš (ed.), *Sudcovské rozhodovanie. Záruky a prekážky spravodlivého procesu*. Bratislava: Wolters Kluwer, pages 216–233.