Tereza Klemensová – Michal Místecký (Ostrava, Czechia)

*The Strange Case of Proper Names: Declension Patterns in Several Czech Onymic Classes**

1. Introduction

In the 2020s, interest of onomasticians in linguistic behaviour of proper names in text seems to have gained momentum, especially in Czech onomastics (cf. CVRČEK et al. 2020, DAVID et al. 2022, DAVID–MÍSTECKÝ 2023, DAVID 2023, MÍSTECKÝ et al. 2024, DAVID–MÍSTECKÝ–DAVIDOVÁ GLOGAROVÁ 2024). So far, most of the studies have focused on case distributions, i.e., on the frequencies of grammatical cases in various types of text. It has already been proved that such research may inform about hidden patterns in texts which deviate from the mainstream narrative one way or another, e.g., in articles from alternative/anti-establishment media (JANDA et al. 2022). A road not taken yet is employing this methodology in diagnosing neurological and psychiatric patients, even though, for instance, in case of so-called voice-hearers, the use of names is one of the distinguishing features regarding the level of agency a voice is endowed with. The fact that a voice a patient hears in their head has a name is thus an important marker of the severity of their medical condition (cf. COLLINS et al. 2023). However, before such research can be undertaken, it is vital to investigate typical case distributions of proper names in what may be considered more neutral context (and hereinafter referred to as general language), this being the primary goal of this paper.

There are in total seven grammatical cases in Czech: the nominative (nom), genitive (gen), dative (dat), accusative (acc), vocative (voc), locative (loc), and instrumental (inst). The main function of the case is building up syntactic relations within a sentence (HAVRÁNEK–JEDLIČKA 1960, ŠTÍCHA et al. 2013). There are cases for which this syntactic function seems to be primary (e.g., the nominative and accusative) and those that in addition to it, possess several semantic traits as well (e.g., the dative, locative, and instrumental). The second group primarily takes the syntactic position of adverbials. The vocative forms a syntactic structure of its own and is thus excluded from the structure of the sentence it is formally part of. The syntactic/semantic division, which stems from the conception of KURYŁOWICZ (1949), has been commonplace in Czech linguistics (cf. KOPEČNÝ 1962, BAUER–GREPL 1972, GREPL et al. 1995); on the other hand, *Mluvnice češtiny 2* (KOMÁREK et al., 1986) approaches cases

* This paper was supported by the project *Quantitative Onomastics: Starting Points, Concepts, Applications* (provided by the Grant Agency of the Czech Republic – GAČR, no. 22-09310S).

from a structuralist viewpoint, construing their system on the grounds of two basic categories—integration and hierarchization—the latter being divided into several other subcategories.

From the perspective of frequency, mainly syntactic cases tend to be much more numerous than those with mostly semantic orientations; this is in line with the overall tendency of language to use features that form its system (e.g., synsemantic words, such as prepositions and conjunctions) considerably more frequently than those with lexical meaning. However, the distributions may differ for proper names, as their functions are supposed to be more specialized—it is, for instance, expectable that toponyms prominently appear in the genitive and locative, as these are the two cases that follow the spatial prepositions *do* 'to' (+ gen), *z* 'from' (+ gen), and *v* 'in' (+ loc).

Studying case distributions of proper names in general language has, therefore, two primary goals: (1) to determine whether these distributions differ from that of appellatives; and (2) to recognize situations in which proper names are not employed as expected. However, such research appears to be a challenging task, at least from two reasons—first, it is difficult to determine what a general language is (the same goes for a representative corpus), and second, there is a very limited number of texts with proper names tagging (i.e., of those in which proper names are labelled). For these reasons—and also for others, which will be addressed further on—the results presented in this paper need to be seen as tentative only, and as completed-to-be (or thoroughly revisited-to-be) once more onomastically tagged language data is at our disposal. For the time being, we make use of OnomOs, the first corpus of Czech language with tagged proper names that is publicly available, as one of the databases of the Czech National Corpus (cf. MÍSTECKÝ et al. 2024). OnomOs contains more than 250,000 tokens and consists of articles of the periodical of Právo (called Rudé právo from 1920 to 1995), which was selected as the oldest Czech newspaper which is still in operation and the publication of which has not been interrupted for a considerably long time (Rudé právo was not officially published between 21 October 1938 and 15 May 1945 only). Being aware of the fact that the results we present are thus genre-specific, we still attempt to address the following questions:

1) Are there statistically significant differences between the case distributions of the selected classes of proper names and that of appellatives?

2) How are the classes of proper names clustered in terms of their case distributions? What is the position of appellatives in such clustering? Given the long-time onomastic search for terminological clarity, does the division of proper names into anthroponyms, toponyms, and

chrematonyms (hereinafter referred to as ACT classification) holds good as concerns the morphological behaviour of proper names?

2. Methods and Material

OnomOs, the corpus we employ in our analyses, is onomastically tagged using the NER (= named-entity recognition) application NameTag 2 (STRAKOVÁ–STRAKA–HAJIČ 2019). In this application, proper names are conceived rather broadly—they include dates, addresses, weekdays, and months. To harmonize the classification used by the software with that employed in mainstream onomastics, the authors of the corpus dropped several categories and merged others in order for the new classes to reflect the current state of debate on onomastic terminology (cf. DAVID–KLEMENSOVÁ–MÍSTECKÝ 2021). Therefore, there are two types of classes in OnomOs—the first-order ones, which are anthroponyms (names of people; A), toponyms (names of places; T), and chrematonyms (names of human-made things; C); and the second-order ones, which provide more information on what the given name names (e.g., AF – first names, TT – territory names, CN – names of periodicals). There are 23 categories in total: four toponym types, five anthroponym types, and fourteen types of chrematonyms. For the full list of the classes, see MÍSTECKÝ et al. 2024.

In our research, we only analyse the case distributions of those categories that have at least 200 occurrences in the corpus. These 200 occurrences need to be morphologically tagged; if the unsuccessful tagging leads to the lowering of the number of occurrences, such name class is excluded from the research. Finally, there are 11 classes that meet our conditions: AF (first names, e.g., *John*), AS (surnames, e.g., *Smith*), AI (inhabitant names, e.g, *an American*); TS (settlement names, e.g., *Prague*), TT (territorial names, e.g., *Czechia*); CA (art products, e.g., *Harry Potter and the Order of the Phoenix*), CC (conferences, contests, and events, e.g., *Champions League*), CF (companies, e.g., *Siemens*), CI (cultural and educational institutions, e.g., *National Theatre*), CN (periodicals, e.g., *Financial Times*), and CP (politics, e.g., *Republican Party*). Concerning the appellatives, we work with a sample of 1,000 non-proprial nouns and adjectives that was selected randomly from OnomOs (corpus query: [pos="A|N"] !within<ne type=".*" />; there are 996 relevant rates of occurrence, as four were not assigned a case). We opted for these two parts of speech as most proper names are either nouns, or adjectives (e.g., *National Theatre*). It is to be noted that multiple-word names are treated as distinct occurrences (*Czech Republic* thus counts as two occurrences of TT); this approach considerably simplifies the process of frequency relativization and makes it possible to compare proper names with appellatives (see Table 1). Last

but not least, the vocative was excluded from the analysis, as it scores very low in many cases and is mostly tagged as such incorrectly.

APEL			**AS**		
case	**fq.**	**relfq.**	**case**	**fq.**	**relfq.**
nom	230	0.23	nom	2,330	0.65
gen	311	0.31	gen	592	0.17
dat	39	0.04	dat	117	0.03
acc	229	0.23	acc	255	0.07
loc	121	0.12	loc	43	0.01
inst	66	0.07	inst	241	0.07
χ^2			777.613		
p-value			min		
significance			YES		
Cramér's V			0.412		

Table 1. *An example of the statistical analysis of the case distributions of two classes: the appellatives (APEL) and surnames (AS).*[1]

Our analysis follows these steps. First, we compare the case distribution of a given proper name class with that of the appellatives; on the grounds of chi-square testing, we arrive at the conclusion whether the two distributions statistically significantly differ. We set the level of significance to be 0.05. This means that if the chi-square statistic corresponds to a value of probability lower than 0.05, there is less than 5% probability that the difference was caused by mere chance—in such a situation, we reject the null hypothesis and consider the case and the proper name class to be associated (cf. LOWRY 1998–2023). We then measure the strength of this association using Cramér's V. Cramér's V values range from 0 to 1. The higher the value of this effect-size measure, the more pronounced the link between the case and name class is. An example of the whole procedure is presented in Table 1.

[1] The abbreviations "fq." and "relfq." stand for "frequency" and "relative frequency" respectively. Under the present research conditions, the value of chi-square test ($\chi 2$) that corresponds to the 5% probability is 11.07; in this particular comparison, it is obviously so high that the corresponding value of probability (p-value) is extremely low ("min"). The result is thus statistically significant at the 0.05 level of significance. Moreover, Cramér's V scores high, too, which indicates a relatively strong association between the word class (APEL vs. AS) and the grammatical cases they give preference to.

Next, we investigate the relations among the classes, utilizing the methodology of hierarchical clustering. Hierarchical clustering is a step-by-step procedure leading to the construction of a dendrogram, a tree diagram that captures similarities/dissimilarities in researched data on the grounds of distances between them. In our analysis, the variables on the grounds of which the distances are calculated are the relativized frequencies of the cases for the individual classes. We use the Euclidean distance formula for a six-dimensional space (as there are six cases to be studied), and subsequently apply the average linkage method, since it seems to avoid certain drawbacks stemming from simple and complete linkage (EMMENDORFER–CANUTO 2021). There are two dendrograms produced: the first, taking into account the proper name classes and the appellatives, and the second, considering solely the proper name classes.

code	**proper name class**	**χ^2**	**p-value**	**sig.?**	**Cramér's V**	**association**
AF	first names	577.32	min	YES	0.475	relatively strong
AS	surnames	777.613	min	YES	0.412	relatively strong
TS	settlements	480.593	min	YES	0.409	relatively strong
CF	companies	276.666	min	YES	0.403	relatively strong
CN	periodicals	193.017	min	YES	0.383	moderate
AI	inhabitants	154.382	min	YES	0.344	moderate
CI	cultural and educational institutions	208.664	min	YES	0.303	moderate
TT	territories	282.593	min	YES	0.302	moderate
CP	politics	185.568	min	YES	0.265	moderate
CC	conferences, contests and events	60.09	min	YES	0.211	moderate
CA	art products	40.174	min	YES	0.167	weak

Table 2. *Results of the statistical tests of the case distributions of the analysed proper name classes (ordered according to the descending values of Cramér's V). The abbreviation "sig." stands for "significance".*

3. Results

The results of the first analysis are listed in Table 2. All the case distributions of the proper name classes differ from that of the appellatives in a statistically significant manner. However, the degree of association of the case and the class, which is expressed by the values of Cramér's V, is varied. The amount of association is interpreted according to REA–PARKER (1992). There are four proper name classes achieving the level of relatively strong association: first names, surnames, settlement names, and names of companies. The case distributions of these are presented in full in Table 3. The weak association of the two categories in the CA–APEL comparison may be attributed to similarities in their case distributions, notably as concerns the frequency of the accusative.

There are two distinct motivations behind the onymic classes that scored high in Cramér's V. The first reason is the dominance of the nominative over the other cases, which is visible in first names (AF), surnames (AS), and, to a slightly lower extent, company names (CF). This morphological feature may be explained by using the first case of AF/AS in rewriting direct speeches of the name bearers (see example 1) or, more broadly, by a general tendency of newspapers to explicitly mention "doers" in order to hold them accountable for their actions (see example 2). The second tendency is represented by settlement names (TS), which, expectedly enough, manifest a soaring frequency of the locative. The reason why territorial names (TT) do not join TS is that they show more levelled occurrences of the locative with the genitive (31% and 32%, as compared to TS's 47% and 27%); this may be ascribed to two factors: (1) TTs tend to be names of countries, which brings them slightly closer to chrematonyms (for which the genitive is one of the typical cases; see MÍSTECKÝ et al., 2024); (2) territories tend to be further divided according to compass points and there is a higher need than in case of TSs to depict various movements from/to them (both the factors combine in example 3).

> [1] Podle Zemana Úřad vyšetřování ČR oficiálně v lednu konstatoval, že daňové úniky jen v případech lehkých topných olejů přesáhly 20 miliard korun. Klaus to označil za nehorázné tvrzení a Zemana nařkl, že v životě nedržel v ruce statistickou ročenku. (*Právo*, 13 May 1996)
>
> According to Zeman [= that-time leader of the Czech Social Democratic Party], the Czech Investigation Office officially stated in January that tax evasion in light fuel oil cases alone exceeded 20 billion Czech crowns. Klaus [= that-time prime minister of Czechia] called this an outrageous claim and accused Zeman of never having held a statistical yearbook in his hands in his life.

[2] Když prezidenti Clinton a Jelcin tento dokument podepsali, otázka východního rozšíření NATO ještě nastolena nebyla. (*Právo*, 13 May 1996)

When Presidents Clinton and Yeltsin signed the document, the question of NATO's eastern enlargement had not yet been raised.

[3] Na pozvání Národního shromáždění přijede dnes do Československa čtrnáctičlenná parlamentní delegace Spojených států brazilských, složená z nejvýznačnějších činitelů brazilských politických stran. (*Rudé právo*, 8 May 1956)

At the invitation of the National Assembly, a 14-member parliamentary delegation of the United States of Brazil, composed of the most prominent members of Brazilian political parties, will arrive in Czechoslovakia today.

	AF		AS		TS		CF	
case	fq.	relfq.	fq.	relfq.	fq.	relfq.	fq.	relfq.
nom	1,041	0.66	2,330	0.65	300	0.16	392	0.56
gen	279	0.18	592	0.17	504	0.27	230	0.33
dat	43	0.03	117	0.03	30	0.02	10	0.01
inst	89	0.06	255	0.07	78	0.04	26	0.04
loc	13	0.01	43	0.01	877	0.47	32	0.05
inst	103	0.07	241	0.07	92	0.05	14	0.02

Table 3. *The case distributions of the onymic classes with V > 0.4, which we perceive as a relatively strong association.*

Next, we present the results of the hierarchical clustering (see Figures 1 and 2). Let us adopt the right–left reading of the dendrograms, which focuses on the closest links first. Going top-down, we see a tight connection between APEL and CA, CP and CC, AS and AF, CF and AI, and TT and TS. Some of these pairs have already been pointed at—AS and AF are connected due to their extensive appearance in the nominative, and toponyms stick together owing to their numerous occurrences in the locative. The connection between APEL and CA (= names of art products) is established by their shared tendency to occur in the accusative (see example 4), which is rather underrepresented in the other name classes. The CF–AI link is very close to that of AS and AF, but the dominance of the nominative is not as striking as in case of the latter; its high frequencies in case of CF are linked to the use of the company names in the citation nominative (e.g., "rescue workers of the concern Jihomoravské.NOM

doly.NOM"; cf. DANEŠ 2009). Last but not least, the CP–CC connection may be explained by the genitive being their dominant grammatical form.

Moving leftwards, we observe further grouping. The CP–CC couple is joined by CI, the nominative and genitive frequencies of which are more balanced. Quite expectedly, the CF–AI group merges with the AS–AF one, forming thus a primarily anthroponymical cluster. At the next level, the CP–CC–CI hyperclass joins the APEL–CA pair, to which CN is added; the special position of CN is caused by its unexpectedly high frequencies of the dative, which is attributable to the newspaper being the addressee/recipient of various statements (see example 5). At this moment, the groups are as closest to the traditional ACT division as they may possibly get—the first cluster (seen top-down) is formed by most chrematonyms and appellatives (APEL, CA, CP, CC, CI, and CN), the second one comprises mostly anthroponyms (AS, AF, CF, AI), and the third one includes toponyms solely (TT, TS). At the leftmost position, the first and the second clusters merge to create a superclass of all-but-toponyms names, with the placenames keeping their idiosyncratic position.

> [4] Po úspěchu „Maryši" obrácena byla pozornost filmových výrobců na hru Aloise Jiráska „Vojnarku", kterou natočí brněnská společnost Terrafilm. (*Rudé právo*, 8 August 1936)
>
> After the success of Maryša, the attention of film producers turned to Alois Jirásek's play Vojnarka, which will be turned into a film by the Brno company Terrafilm.
>
> [5] Jak Právu sdělil ředitel sekretariátu Sdružení automobilového průmyslu Antonín Šípek, celkem je podle Centrálního registru vozidel podchyceno 36 výrobních značek autobusů. (*Právo*, 13 May 1996)
>
> Antonín Šípek, director of the secretariat of the Association of the Automotive Industry, told Právo that according to the Central Vehicle Register, a total of 36 brands of buses are registered.

The importance of the cases is also reflected in their dendrogram, which is part of Figure 1 as well (see the top of it). Adopting a bottom-up reading of the dendrogram, we see two pairs: the nominative–genitive one, the occurrences of which are decisive for the grouping of proper names, and the one of the dative and instrumental, which share very low frequencies. It is to this cluster that the accusative and locative are one by one joined. It is to be further investigated whether the prevalence of nominative and genitive forms in most onymic classes is their general feature, or whether it is attributable to the genre of newspaper articles.

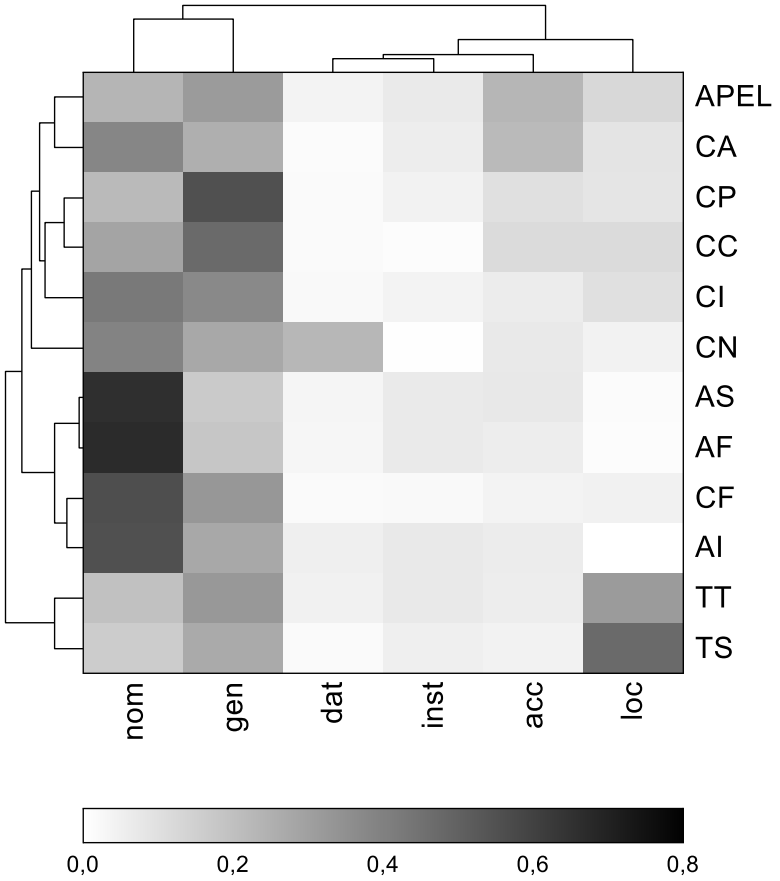

Figure 1. *The dendrogram depicting the hierarchical clustering of the onymic classes and the appellatives (APEL) on the grounds of the relative frequencies of the grammatical cases. The topmost graphic captures the hierarchical clustering of the cases. Concerning the heatmap, the darker the shade, the more frequent a particular case is in the occurrences of the particular class.*

Exclusion of the appellatives from analysis has brought about some changes to the clusters. Adopting the right–left reading again, we see the pairs of AS and AF, CF and AI, CI and CA, CP and CC, and TT and TS. The only novice here is the CI–CA link, which is a product of the lack of partners the accusative-occurring CA could join—the connection between CI and CA thus being their balanced frequencies of the nominative and genitive. At the next level, anthroponyms and CF form the same cluster as in Figure 1, but CN joins the

CI–CA couple, the main reason being again the nominative–genitive balance. The biggest change comes with toponyms, which are joined to the CP–CC couple, probably on the grounds of prevalence of the genitive over the nominative. In this second grouping, the chrematonyms are thus more divided than when the appellatives were part of the picture. Slight modifications are also detectable in the topmost case dendrogram—while the nominative–genitive pair is intact, in the second cluster, the accusative is first linked to the instrumental, and the dative comes at the second level.

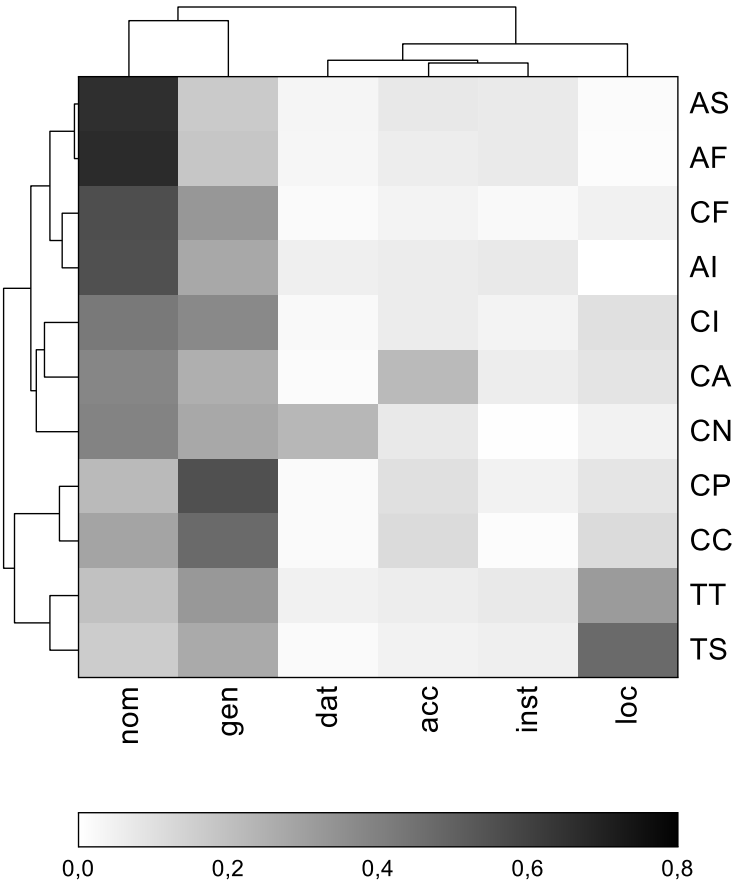

Figure 2. *The dendrogram depicting the hierarchical clustering of the onymic classes on the grounds of the relative frequencies of the grammatical cases. The topmost graphic captures the hierarchical clustering of the cases. Concerning the heatmap, the darker the shade, the more frequent a particular case is in the occurrences of the particular class.*

4. Conclusion

As regards the two questions we asked in the Introduction, the first one may be answered in the affirmative—there are statistically significant differences between the case distribution of the appellatives and the distributions of individual proper name classes. The difference seems to be the most relevant for first names, surnames, settlement names, and company names, which points at the anthroponyms being the most distinctive group of proper names in terms of morphology. Settlement names, on the other hand, manifest high frequencies of the genitive and locative, which brings them closer to territorial names.

Concerning the second question, and especially the sense of the ACT division, the cluster analyses confirm the tight connections among/between various types of anthroponyms and toponyms, but as for chrematonyms, the structures are rather cluttered. As has been already mentioned, company names morphologically behave like anthroponyms, CN (periodicals) manifests a lot of dative forms, CA (art products) does appear in the accusative, and CI's (cultural/educational institutions) proportions of the nominative and genitive are almost equal. The CP–CC (politics; conferences, contests, and events) link is strong, but in the second dendrogram, it is connected to the toponymical cluster. These findings make it difficult to sketch shared morphological characteristics for chrematonyms. More research—for instance in the domain of syntactic functions—is thus needed to understand the grammatical workings of this first-order proper name class. From the perspective of grammatical cases, the dendrograms declare the central position of the nominative–genitive cluster in the frequency distributions of the studied proper names, with several notable exceptions (cf. the accusative in case of CA, art products). The interplay between the nominative and genitive rates of occurrence seems to be one of the most important factors behind both the clustering outputs.

Defining grammatical links between proper name groups may provide onomastics with new insights. Namely, it may lead to (1) perfecting the tagging procedure and result in onomastic taggers working with a lower error rate; (2) revisiting the position of chrematonyms in the system and designing their new classification, which would be more in line with their morphological behaviour (e.g., in terms of their distinction from appellatives); (3) detecting language products in which names behave in a marked way: political discourse, fake news detection (e.g. not using territorial names in the locative, see DAVID–MÍSTECKÝ 2025), advertisements, literature studies, or neurological patients' language (e.g. foregrounding and backgrounding of certain anthroponyms via the nominative/accusative distinction).

There are certain limitations to the research presented in the paper. First, the selection of the appellatives was random, but it cannot be ruled out that it is

skewed in one direction or another. Second, there is an error rate of the morphological tagger used by the Czech National Corpus; not all the instances may thus have been labelled correctly. And third, grammatical cases have multiple functions, the morphological closeness of the names thus always needs to be confronted with the actual examples from the corpus.

Moreover, as mentioned already in the Introduction, it is advisable the research be replicated on a more complex language material, as the present conclusions are limited genre-wise. Bringing syntax and lexicology in the game, for instance via collocation analyses, may help construct a comprehensive picture of language behaviour of proper names and shed more light on the important philosophical issue of their meaning/sense as well.

References

BAUER, JAROSLAV–GREPL, MIROSLAV 1972. *Skladba spisovné češtiny*. [The syntax of standard Czech.] Prague, SPN.

COLLINS, LUKE–BREZINA, VACLAV–DEMJÉN, ZSÓFIA–SEMINO, ELENA–WOODS, ANGELA 2023. Corpus linguistics and clinical psychology: Investigating personification in first-person accounts of voice-hearing. *International Journal of Corpus Linguistics 28:* 28–59.

CVRČEK, VÁCLAV–LAUBEOVÁ, ZUZANA–LUKEŠ, DAVID–POUKAROVÁ, PETRA–ŘEHOŘKOVÁ, ANNA–ZASINA, ADRIAN JAN 2020. *Registry v češtině*. [Registers in Czech.] Prague, NLN.

DANEŠ, FRANTIŠEK 2009. *Kultura a struktura českého jazyka*. [The culture and structure of the Czech language.] Prague, Karolinum.

DAVID, JAROSLAV 2023. Czech place names and their exonyms in parallel corpus – between preserving the original form and adaptation. *Onomastica 67:* 205–220.

DAVID, JAROSLAV–KLEMENSOVÁ, TEREZA–MÍSTECKÝ, MICHAL 2021. Věda mnoha jmen: onomastické termíny v publicistice Českého národního korpusu. [Science of many names: Onomastic terms in the opinion journalism of the Czech National Corpus.] *Jazykovedný časopis 72:* 114–123.

DAVID, JAROSLAV–KLEMENSOVÁ, TEREZA–MÍSTECKÝ, MICHAL–DAVIDOVÁ GLOGAROVÁ, JANA–MÁDROVÁ, JARMILA–BŘEZINOVÁ, KRISTÝNA–KOVÁŘOVÁ, KRISTÝNA–MÍSTECKÁ, DENISA–RECLIK, AGATA D.–NEUWIRTHOVÁ, KLÁRA–ČERNÍKOVÁ, TEREZA 2022. *Od etymologie ke*

krajině: Onomastika pro 21. století. [From etymology to landscape: Onomastics for the 21st century.] Brno, Host.

DAVID, JAROSLAV–MÍSTECKÝ, MICHAL 2023. Prolegomena ke kvantitativní onomastice. [Prolegomena to quantitative onomastics.] *Acta onomastica 64:* 301–320.

DAVID, JAROSLAV–MÍSTECKÝ, MICHAL 2025. Locatives Lost: Case Distributions of Toponyms in Anti-Establishment Media. *Onoma* [accepted for publication].

DAVID, JAROSLAV–MÍSTECKÝ, MICHAL–DAVIDOVÁ GLOGAROVÁ, JANA 2024. Kvantitativněonomastická analýza Haškova Švejka a jeho tří parafrází [Quantitative-onomastic analysis of Hašek's Švejk and its three paraphrases.] *Slovo a slovesnost 85:* 199–216.

EMMENDORFER, LEONARDO RAMOS–CANUTO, ANNA MAGALY DE PAULA 2021. A generalized average linkage criterion for Hierarchical Agglomerative Clustering. *Applied Soft Computing 100:* article 106990.

GREPL, MIROSLAV–KARLÍK, PETR–NEKULA, MAREK–RUSÍNOVÁ, ZDENKA 1995. *Příruční mluvnice češtiny*. [A concise grammar of Czech.] Prague, NLN.

HAVRÁNEK, BOHUSLAV–JEDLIČKA, ALOIS 1960. *Česká mluvnice*. [Czech grammar.] Prague, SPN.

JANDA, LAURA A.–FIDLER, MASAKO–CVRČEK, VÁCLAV–OBUKHOVA, ANNA 2022. The case for case in Putin's speeches. *Russian Linguistics 47:* 15–40.

KOMÁREK, MIROSLAV–KOŘENSKÝ, JAROSLAV–PETR, JAN–VESELKOVÁ, JARMILA et al. 1986. *Mluvnice češtiny 2*. [Czech grammar 2.] Prague, Academia.

KOPEČNÝ, FRANTIŠEK 1962. *Základy české skladby*. [Fundamentals of Czech syntax.] Prague, SPN.

KURYŁOWICZ, JERZY 1949. Le problème du classement des cas. *Bulletin de la Société Polonaise de Linguistique 9:* 20–43.

LOWRY, RICHARD 1998–2023. *Concepts & Applications of Inferential Statistics*. http://vassarstats.net/textbook/ (accessed 2025-10-04).

MÍSTECKÝ, MICHAL–DAVID, JAROSLAV–DAVIDOVÁ GLOGAROVÁ, JANA–KLEMENSOVÁ, TEREZA 2024. OnomOs: principy a příklady aplikací. [OnomOs: principles and examples of applications.] *Časopis pro moderní filologii 106:* 44–59.

REA, LOUIS M.–PARKER, RICHARD A. 1992. *Designing and conducting survey research*. San Francisco, Jossey-Boss.

STRAKOVÁ, JANA–STRAKA, MILAN–HAJIČ, JAN 2019. Neural Architectures for Nested NER through Linearization. In: KORHONEN, ANNA–TRAUM, DAVID–MÀRQUEZ, LLUÍS eds. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Stroudsburg, Association for Computational Linguistics. 5326–5331.

ŠTÍCHA, JAN–VONDRÁČEK, MILOSLAV–KOLÁŘOVÁ, IVANA–HOFFMANNOVÁ, JANA–BÍLKOVÁ, JANA–SVOBODOVÁ, IVANA 2013. *Akademická gramatika spisovné češtiny*. [Academic grammar of Standard Czech.] Prague, Academia.

Abstract

The paper analyses grammatical case distributions of appellatives and eleven onymic classes (first names, surnames, inhabitant names, settlement names, territorial names, art product names, names of conferences, contests, and events, names of companies, names of cultural and educational institutions, names of periodicals, and names connected to politics). The frequencies of the cases are taken from OnomOs, a Czech corpus with tagged proper names and comprising articles from the left-leaning periodical of (Rudé) Právo that span the period of 1920s to 2010s. Only the proper name classes with more than 200 morphologically tagged occurrences are taken into account. The first analysis has shown that all the proper name case distributions are statistically significantly different from that of the appellatives; first names, surnames, settlement names, and company names were identified as the most deviating structures. The second analysis has pointed at similarities among anthroponymical and toponymical classes, but discovered morphological diversity in case of chrematonyms. The research is to be supplemented with investigations focused on syntactic functions and lexical aspects of names, e.g., on collocations.

Keywords: declension, case, grammar, proper name, corpus