sciendo
DOI 10.2478/jazcas-2023-0027

PROVERBS IN CONTEMPORARY CZECH. CORPUS PROBE INTO WRITTEN TEXTS

MARIE KOPŘIVOVÁ[1] – KATEŘINA ŠICHOVÁ[2]
[1] Institute of Czech and Deaf Studies, Faculty of Arts, Charles University, Prague, Czech Republic
[2] Bohemicum – Center for Czech Studies, University of Regensburg, Regensburg, Germany

KOPŘIVOVÁ, Marie – ŠICHOVÁ, Kateřina: Proverbs in Contemporary Czech. Corpus Probe into Written Texts. Journal of Linguistics, 2023, Vol. 74, No 1, pp. 92 – 99.

Abstract: The paper deals with the possibility of creating a paremiological optimum for students of Czech as a foreign language. The selection of proverbs should reflect the frequency, familiarity with and use of proverbs. The study focuses on the most frequent proverbs in written Czech, using contemporary idiomatically annotated corpora. On this basis, our own minimum was created. The paper compares the results with previous studies on the paremiological minima of Czech (Schindler 1993 and Čermák 2003) and shows the intersection of all three minima.

Keywords: paremiological minimum/optimum, idiom annotation, Czech proverbs, corpus

1 INTRODUCTION

Paremiology has enjoyed scholarly interest for several centuries, and we already have the Book of Proverbs in the Bible, although it contains a somewhat different kind of text than we are accustomed to understand by proverbs today; its content is identical to today's proverbs in that it gives advice on how to live well. The Adagia of Erasmus of Rotterdam from 1500 is considered one of the oldest collections of European phraseology, but there are also extensive collections in Czech, e.g. by Comenius (first edition 1849), Čelakovský (1852) and Flajšhans (1911 and 1913, new edition 2013). All these authors tried to collect as many proverbs as possible, searching them mostly in literature and sometimes translating them from other languages. From today's point of view, they could be called an attempt at phraseological maximum.

Proverbs have been studied from a professional point of view not only in linguistics, which deals with them as special linguistic units and classifies them as idioms of sentence nature (cf. Čermák et al. 2009, hereafter SČFI), but also, for example, in folkloristics, which views them as a condensed experience (e.g.

Bittnerová – Schindler 2003), and phraseodidactics, which focuses on practicing them (cf. Šemelík – Šichová 2016, 2017; Jesenšek 2013). The continued interest in this topic is also evidenced by a number of other activities focused on proverb research, such as the Proverbium journal,[1] the international conference Interdisciplinary Colloquium on Proverbs,[2] and a web database capturing proverbs from different languages.[3]

There are also many popularising book and web collections or teaching materials.[4] Proverbs have a special place in teaching Czech as a mother tongue and as a foreign language and knowledge of their form and meaning is tested in examinations.

The problem is that proverbs are linguistic units of several words and as such are not very frequent in texts. In addition, some proverbs contain archaic realities and may therefore be considered by students to be outdated and no longer in use today. For this purpose, we no longer need the most extensive collection of proverbs, where we do not find information about their use, but an idea of which proverbs are the ones known and used.

As far as the definition of proverb is concerned, we conceive of it rather broadly and lean towards the definition of the paremiologist W. Mieder: "Proverb is commonly thought of as a phrase, saying, sentence, statement, or expression of the folk which contains above all wisdom, truth, morals, experience, lessons, and advice concerning life and which has been handed down from generation to generation." (Mieder 1993, p. 24).

Another important characteristic of proverbs is their stability in sentence form and the fact that they can be understood without adaptation to context (Schindler 1996, p. 265). They share this property with quotations. Therefore, they appear in texts as headings, summaries or comments on certain events, and they tend to be introduced by special introducers or formulas. (cf. Čermák 2007, pp. 549–568).

2 PREVIOUS RESEARCH

Mainly with regard to lexicography (problematic definition of paremiological units and traditionalisation of obsolete or unusual units) and phraseodidactics (the need to select appropriate units for the purpose of teaching a given language as a foreign/second language), the "maximalist demand" and attempts to document all proverbs of a given language, including variants, are abandoned (Ďurčo 2015b, p. 41). There are efforts to produce more quantitatively limited sets, e.g. to filter a certain core of the paremiological vocabulary of a given language.

[1] https://naklada.ffos.hr/casopisi/

[2] https://www.folklore.ee/rl/fo/koostoo/tavira/proceedings.htm

[3] http://www.sprichwort-plattform.org/

[4] e.g. https://www.presto.cz/cz/ceska-prislovi, https://www.knihovnahk.cz/files/tinymce/Prislovi.pdf

The work of the Russian paremiologist Grigoriy Permyakov was groundbreaking in this respect. Permyakov's aim (1971, cited in Mokienko 2012) was to compile the "paremiological fund" of Russian, i.e., to identify the most frequent proverbs. From these, he then tried to identify the minimum set of proverbs known to speakers, the so-called "paremiological minimum" (for Russian, he put it at 300 proverbs). Since Permyakov's idea, attempts have been made to select proverbs in a scientific way for other languages (e.g. for English Mieder 1994; Haas 2008; for German Baur – Chlosta 1996a, 1996b; Ďurčo 2015a; Grzybek 2012; for Slovak Ďurčo 2015b: 196ff.; for more cf. Mokienko 2012, pp. 79–83 and Ďurčo 2015a), while from the point of view of language teaching it has over time proved more appropriate to speak of the paremiological optimum of the respective language rather than the minimum.

The approaches to obtaining material and the criteria for selecting paremiological units vary (Jelínek et al. 2018), partly conditioned by lexicographic and/or linguistic goals (Hrisztova-Gotthard – Varga 2014): analysis of scientific publications (e.g. Mieder 1994), corpus search (e.g. Steyer 2012), analysis of dictionaries (e.g. Hessky – Ettinger 1997).

2.1 Paremiological minima in Czech

For Czech F. Schindler (1993), following Permajakov's model, compiled a paremiological minimum: in a questionnaire survey, he had individual speakers complete the second part of the proverb and then evaluated the degree of knowledge. His great contribution is the very selection of the proverbs to be tested; he used the above-mentioned collections, especially Čelakovský's, and the glossary of the forthcoming 4th volume of the SČFI (Čermák et al. 2009). The selection of proverbs is made more difficult by the lack of a standard definition. However, Schindler, based on his research, established 99 proverbs as the paremiological minimum (Schindler 1996); other proverbs for which respondents showed lower knowledge are included in the book Czech Proverbs (Bittnerová – Schindler 2003). This research was used by F. Čermák as a starting point to verify the frequency of individual proverbs in the corpus (Čermák 2003); for his research he chose the 100-million representative corpus SYN2005 and determined 100 most frequent proverbs in this corpus.

2.2 Paremiological optimum and the prerequisites for its creation

Nowadays, some linguists argue for a critical view of the minima, e.g. Mokienko states: "It is didactic goals and lexicographical description that determine in each individual case the particular paremiological minima suitable for particular addressees or groups of addressees. A fixed paremiological minimum, i.e., a minimum for all language users, is a purely scientific abstraction if not an outright fiction." However, if the knowledge of proverbs is to be taught among students of Czech as a foreign language, it seems essential to create a paremiological optimum,

i.e., a set of proverbs that are frequent (used in Czech texts) and/or known (actively, passively by speakers of Czech). It is important to distinguish between knowledge of form and knowledge of meaning. However, the latter is difficult to verify, and it is similarly problematic to find out directly from speakers which proverbs they actively use in everyday communication (cf. Grzybek 2012).

The aim of our research is to determine the phraseological optimum for learners of Czech with German as their mother tongue. The starting point would be a proverb pool from which proverbs meeting the relevant criteria (frequency, familiarity, usage) would be selected. These basic criteria can be extended by additional criteria specific to certain user groups, thus obtaining paremiological optima suitable for specific purposes. In our case, we could add as an additional criterion, for example, the existence of an equivalent proverb in German.

Some proverbs in European languages have similar origins (cf. e.g. Tölgyesi 2022), but the approach to their use may be different (cf. Schindler 1996, p. 280), which is reflected in teaching and testing. Passive comprehension is also important for learners of Czech as a foreign language, and thus the frequency with which they encounter proverbs plays a large role here. The frequency of occurrence of proverbs can be quite reliably determined in written Czech using corpora. Therefore, in the first step, we only start from the frequency of proverbs and from a comparison with previous paremiological minima (Schindler 1996; Čermák 2007). We want to determine the consistency between the different approaches.

3 DATA AND METHODOLOGY

Compared to Čermák's study, we used more extensive data, in which the annotation of idioms is additionally present (cf. Hnátková 2002). However, proverbs do not form a separate group within the annotation, but are part of a set of sentence idioms, that must be searched manually to decide which ones are proverbs. In this section, we have verified the existence of proverbs in the Dictionary of Proverbs (Čermák 2013). Currently, a more detailed processing of proverbs is underway using the LEMUR database, which will also serve to better annotate proverb fragments. Unfortunately, this database is not yet available to a satisfactory extent (Hnátková et al. 2017).

It is possible to use more extensive data from a larger time period, which includes the syn_v9 corpus (with texts from 1990–2019). This corpus is heavily dominated by journalism, so it is not suitable for our probe; the frequency of proverbs would reflect their use in journalism. To limit the influence of the journalistic genre, a balanced corpus consisting of 4 representative corpora of written Czech was chosen: SYN2000, SYN2005, SYN2010, SYN2015.

This corpus contains 32% fiction, 40% journalistic and 28% professional literature. The highest number of proverbs is found in journalistic, then in fiction,

and the lowest in professional literature, which includes popular literature. To reduce the influence of journalistic literature, three lists of proverbs were created according to each type of text, and the set of those proverbs that appeared in all three lists is hereafter referred to as paremiological minimum 2023 (PM23).

4 COMPARISON OF PAREMIOLOGICAL MINIMA RESULTS

The 100 most frequent proverbs from PM23 were compared with Schindler's (99 proverbs) and Čermák's minimum (100 proverbs). The higher agreement with Čermák's minimum was confirmed, with 57 identical proverbs, and only 31 with Schindler's; agreement with Čermák or Schindler accounted for 62 proverbs. The intersection of all three was 28 proverbs. It is not clear whether the differences are more due to a temporal shift or to the composition of the data. In any case, we believe that these proverbs could form the core of the paremiological optimum.

They are the following proverbs:

1. *Šaty dělají člověka.* 'Clothes make the man.' 'Fine feathers make fine birds.'
2. *Dvakrát měř, (a) jednou řež.* 'Measure twice, cut once.' 'Look twice before you leap.'
3. *Vlk se nažral a koza zůstala celá.* 'The wolf has eaten and the goat has remained entire.'[5]
4. *Stará láska nerezaví.* 'Old love is never forgotten.'
5. *Bližší košile než kabát.* 'Near my coat, but nearer my skin (shirt).'
6. *Bez práce nejsou koláče.* 'No work, no cake.' 'No cross, no crown.'
7. *Kdo nic nedělá, nic nezkazí.* 'He who does nothing, spoils nothing.'
8. *Sejde z očí, sejde z mysli.* 'Out of sight, out of mind.'
9. *Co na srdci, to na jazyku.* 'What's on the heart is on the tongue.'
10. *Lež má krátké nohy.* 'Lies have short wings (legs).'
11. *Kdo hledá, najde.* 'Search and you shall find.'
12. *Lepší vrabec v hrsti než(li) holub na střeše.* 'Better a sparrow in the hand than a pigeon on the roof.' 'A bird in hand is worth two in the bush.'
13. *Všude dobře, doma nejlíp.* 'There is no place like home.'
14. *I mistr tesař se utne.* 'Even a master carpenter gets it wrong.'
15. *Jablko nepadá daleko od stromu.* 'The apple doesn't fall far from the tree.' 'Like father, like son'. 'He is a chip of the old block.'
16. *Láska hory přenáší.* 'Love moves mountains.'
17. *Vrána k vráně sedá.* 'The crow sits with the crow.'
18. *Tonoucí se stébla chytá.* 'A drowning man clutches at a straw.'

[5] For the expression *the wolf has eaten and the goat has remained entire* we can doubt its classification as a proverb if we follow the formal criterion for proverbs, which is its independence and syntactic uninvolvement in the sentence. This expression is usually part of a sentence with a purposeful meaning: to do something so that the wolf eats and the goat stays entire.

19. *Čert nikdy nespí.* ‘Devil never sleeps.’

20. *Kuj železo, dokud je žhavé.* ‘Strike while the iron is hot.’

21. *Co Čech, to muzikant.* ‘Every Czech is a musician.’

22. *Každý svého štěstí strůjcem.* ‘Each man is the author of his own happiness.’ ‘Life is what you make it.’

23. *Kdo dřív přijde, ten dřív mele.* ‘First come, first served.’

24. *Na hrubý pytel hrubá záplata.* ‘For a coarse sack a coarse patch.’ ‘Meet rudeness with rudeness.’

25. *Ráno moudřejší večera.* ‘Morning is wiser than evening.’

26. *Boží mlýny melou pomalu, ale jistě.* ‘The mills of God grind slowly but surely.’

27. *Na každém šprochu pravdy trochu.* ‘There’s a little truth in every speck.’

28. *Mluviti stříbro, mlčeti zlato.* ‘Silver to speak, gold to keep silent.’

The list is dominated by short proverbs, which generally have less variation and their form is more easily reproduced, this ensures a higher frequency. In Schindler’s research, this feature does not play a role, but it is reflected to some extent in the familiarity with proverbs: speakers remember short proverbs better, while longer ones show higher variability. Thus, the intersection of all three minima results in shorter proverbs.

5 FUTURE RESEARCH

We plan to repeat the selection of proverbs using the more extensive and recent data by adding a representative corpus of syn2020. Further work is needed to check the annotation with respect to its scope: different variants of proverbs are not always included in the annotation, while it is the variability that can be high in longer units and should be included in the overall frequency of proverbs. There may also be deliberate modifications, which may indicate that the author considers the proverb to be so well known that he can actualize it. Next, we focus on the distribution of individual proverbs within the data, as there may be a cumulation of proverb occurrences due to the proverb being used as the title of a book, exhibition, or to indicate some current information; for this case, an average reduced frequency (ARF) would be appropriate.

In further research, we also want to focus on the selection of proverbs in lexicographic processing, in the production of textbooks, on comparison with the German paremiological optimum, and on research on proverbs among native speakers in terms of proverb familiarity and usage.

References

Baur, R. S., and Chlosta, Ch. (1996a). Welche Übung macht den Meister? In Fremdsprache Deutsch, 15, pages 17–24.

Baur, R. S., and Chlosta, Ch. (1996b). Sprichwörter: ein Problem für Fremdsprachenlehrer wie lerner?! In Deutsch als Fremdsprache, 33, pages 91–102.

Bittnerová, D., and Schindler, F. (2003). Česká přísloví. Soudobý stav konce 20. století. Praha: Karolinum.

Čermák, F. (2013). Základní slovník českých přísloví: Výklad a užití. Praha: Nakladatelství Lidové noviny.

Čermák, F. (2007). Frazeologie a idiomatika česká a obecná. Czech and General Phraseology. Praha: Karolinum.

Čermák, F. (2003). Paremiological Minimum of Czech: The Corpus Evidence. In H. Burger – A. Häcki Buhofer – G. Gréciano (eds.): Flut von Texten – Vielfalt der Kulturen. Hohengehren: Schneider Verlag, pages 15–31.

Čermák, F., Holub, J., Blatná, R., and Kopřivová, M. (2009). Slovník české frazeologie a idiomatiky (SČFI), vol. 4. Praha: Leda.

Ďurčo, P. (2015a). Diasystematische Differenzen von Sprichwörtern aus der Sicht der kontrastiven Parömiografie. In P. Ďurčo – K. Steyer – K. Hein: Sprichwörter im Gebrauch. Trnava, pages 121–142.

Ďurčo, P. (2015b). Empirical Research and Paremiological Minimum. In H. Hrisztova-Gotthardt – M. A. Varga (eds.): Introduction to Paremiology. A Comprehensive Guide to Proverb Studies. De Gruyter Open, pages 183–205.

Grzybek, P. (2012). Facetten des parömiologischen Rubik-Würfels Kenntnis ≡ Bekanntheit [⇔ Verwendung ≈ Frequenz] ?!? In K. Steyer (ed.): Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömiologie. Narr Verlag, pages 99–138.

Hnátková, M. (2002). Značkování frazémů a idiomů v Českém národním korpusu s pomocí Slovníku české frazeologie a idiomatiky. Slovo a slovesnost, 63(2), pages 117–126.

Hnátková, M., Jelínek, T., Kopřivová, M., Petkevič, V., Rosen, A., Skoumalová, H., and Vondřička, P. (2017). Eye of a Needle in a Haystack. Multiword Expressions in Czech: Typology and Lexicon. In R. Mitkov (ed.): Computational and Corpus-Based Phraseology: Second International Conference, Europhras 2017. London: UK, November, pages 13–14.

Hrisztova-Gotthard, H., and Varga, M. A. (2014). Introduction to Paremiology. A Comprehensive Guide to Proverb Studies. De Gruyter Open.

Hessky, R., and Ettinger, S. (1997). Deutsche Redewendungen. Ein Wörter- und Übungsbuch für Fortgeschrittene. Tübingen.

Jesenšek, V. (2013). Sprichwortgebrauch heute. Linguistische und sprachdidaktische Überlegungen. In Muttersprache 2, pages 81–98.

Jelínek, T., Kopřivová, M., Petkevič, V., and Skoumalová, H. (2018). Variabilita českých frazémů v úzu. Časopis pro moderní filologii, 100(2), pages 151–175.

Křen, M., Cvrček, V., Henyš, J., Hnátková, M., Jelínek, T., Kocek, J., Kováříková, D., Křivan, J., Milička, J., Petkevič, V., Procházka, P., Skoumalová, H., Šindlerová, J., and Škrabal, M. (2021). Korpus SYN, verze 9 z 5. 12. 2021. Ústav Českého národního korpusu FF UK, Praha. Accessible at: https://www.korpus.cz.

Mieder, W. (1994). Paremiological Minimum and Cultural Literacy. In W. Mieder (ed.): Wise Words. Essays on the Proverb. Garland Publishing, pages 297–316.

Mieder, W. (1993): Proverbs are Never Out of Season: Popular Wisdom in the Modern Age. Oxford University Press.

Mokienko, V. M. (2012). Russisches parömiologisches Minimum: Theorie oder Praxis? In K. Steyer (ed.): Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömilogie. Tübingen: Narr Verlag, pages 79–98.

Schindler, F. (1996). Sociolingvistické, paremiologické a paremiografické výsledky empirického výzkumu znalosti přísloví. Slovo a slovesnost, 57(4), pages 264–282.

Schindler, F. (1993). Das Sprichwort im heutigen Tschechischen. Empirische Untersuchung und semantische Beschreibung. Otto Sagner, München.

Steyer, K. (2012). Sprichwortstatus, Frequenz, Musterbildung. Parömiologische Fragen im Lichte korpusmethodischer Empirie. In K. Steyer (ed.): Sprichwörter multilingual. Theoretische, empirische und angewandte Aspekte der modernen Parömilogie. Narr Verlag, pages 287–314.

Šichová, K., and Šemelík, M. (2016). Was nicht ist, kann noch werden. Zur Parömiodidaktik im Tschechisch-als-Fremdsprache-Unterricht. Teil 1: Eine exemplarische Lehrwerk- und Wörterbuchanalyse. In Beiträge zur Fremdsprachenvermittlung, 57, pages 61–104.

Šichová, K., and Šemelík, M. (2017). Was nicht ist, kann noch werden. Zur Parömiodidaktik im Tschechisch-als-Fremdsprache-Unterricht. Teil 2: Aufgaben und Übungen. In Beiträge zur Fremdsprachenvermittlung, 59, pages 37–76.

Tölgyesi, T. (2022). Deutsch im interlingualen und interkulturellen Vergleich. In S. Szatzker – A. Szilágyi-Kósa (eds.): Hamburg: Dr. Kovač, pages 157–182.