Mudra's Upper Sorbian-Czech dictionary – what can be done about this lexicographic "posthumous child"?

Michal Škrabal[1], Katja Brankačkec[2]

[1] Charles University, Institute of the Czech National Corpus,
Panská 7, 110 00 Praha 1, Czech Republic
[2] Institute of Slavonic Studies of the Czech Academy of Sciences,
Valentinská 1, 110 00 Praha 1, Czech Republic
E-mail: michal.skrabal@ff.cuni.cz, brankatschk@slu.cas.cz

Abstract

Jiří Mudra, among his numerous selfless activities, was a Czech doyen of Sorbian studies. He had been working for decades on an Upper Sorbian-Czech dictionary but, unfortunately, had not finished his work on it at the time of his death. Presently, we are considering completing Mudra's project. The material collected by Mudra is undoubtedly valuable for us, providing us with a launchpad for further work; still, it is necessary to challenge it with the current data and a modern lexicographic approach. The paper presents the proposed individual methods aimed at finishing the main body of the dictionary.

Every lexicographer works with the data and tools available in his or her time – and Mudra was certainly no exception. There is, therefore, no reason to maintain exaggerated reverence towards his dataset where it is in apparent conflict with the current language reality. The aim is not to foster Mudra's cult, but to acknowledge his admirable initiative and enthusiasm. The best way to do so is to complete his dictionary with all the possibilities currently offered to us and make it available – as the first academic dictionary in this language combination – to Czech users.

Keywords: Upper Sorbian-Czech dictionary; completion; Jiří Mudra

1. Introduction

In recent years, there has only been a small amount of literature on Upper Sorbian (US)[1] lexicography from the methodological perspective (Itoya, 2013; Šěrakowa, 2009; Pohončowa, 2008; Pohončowa & Šołćina, 2007), let alone on *digital* lexicography (Bartels et al., 2021). The following two sections summarise the leading publications on US lexicographic works from the past 30 years, including Jiří Mudra's unfinished US-Czech dictionary. The plan to complete this dictionary is then revealed in Chapter 2. The final chapter answers the question asked in the title of our paper.

---

[1] As our interest lies in US, we omit Lower Sorbian (LS) in our paper, referring only to these sources: Leszcyński (2013), Szpila (2014) and Bartels et al. (forthcoming).

1.1 US lexicography and dictionaries nowadays

Sorbian lexicography only stands on the threshold of the digital-born era. In the most recent paper on it (Bartels et al., 2021), we even find the categoric statement that "there is not even one Sorbian dictionary based on a systematic and extensive analysis of written texts". The authors describe their plans and their initial experience with the first project of digital lexicography targeting Sorbian. The project aims to identify neologisms and changes in the usage of the Upper and Lower Sorbian lexicon. For this purpose, a new corpus of Sorbian texts published by the Domowina publishing house (which covers an overwhelming majority of all officially published Sorbian texts) starting from 2019 is being created.

However, there are some dictionaries available digitally – both in Upper and Lower Sorbian. The most important of them is "Soblex",[2] which aggregates material from the following printed dictionaries and other language tools:

- the 5th edition of the US-German spelling dictionary (Völkel, 2005), initially prepared by P. Völkel (1931–1997). The first edition comes from 1970; further, revised editions appeared in 1976, 1979, and a larger version with spelling rules in 1981. The 5th edition, a revised and enlarged version with the new spelling rules, was prepared by T. Meškank (2005), also with a CD-ROM version (2008). It was edited again in 2014, and is now being processed in a modern way for a wholly revised 8th edition based on modern lexicographical methods that is envisaged to be published during the 2030s. Meanwhile, there also seems to be an internal review of the 7th edition, planned for 2022 (Bartels et al., 2021);
- the German-US dictionary of neologisms (Jenč et al., 2006). This latest printed dictionary (not considering reissues) is designed to extend the two-volume German-US dictionary published earlier (Jentsch et al., 1989, 1991). These older volumes should soon expand the Soblex infrastructure;
- the dictionary for native speakers in US schools (Hajduk-Veljkovićowa, 2017);
- the dictionary of Sorbian names (Meškank, 2017);
- various terminological dictionaries for US schools (e.g. Korjeńk, 1995; Mehrowa & Pawlikowa, 1996);
- the tool integrating US into the Linux OS (including US interface, spellchecker, dictionary of synonyms, among others) developed by E. Werner at the Leipzig University;[3]
- the morphological tool "SorbOrto", developed by G. Nagora and G. Müller;
- the series produced for the US broadcasting programme "Rěčne kućiki" ('Language columns');[4]

- the database of geographic exonyms prepared by the Sorbian Institute and the Witaj Language Centre;[5]
- the most recent add-on is a translating system "sotra"[6] ('sister') in both US-German-US directions. It works on the basis of the statistical translation software "Moses" that gradually learns from a parallel corpus prepared by the Witaj Language Centre.

There is also a set of older printed dictionaries available in digital form, e.g., the online version of the two-volume German-US dictionary (Jentsch et al., 1989, 1991) on the portal hornjoserbsce.de,[7] covering more than 36,000 entries. This website is closely linked to the Soblex dictionary and will soon become a part of it (Bejmak et al., forthcoming).

The older dictionaries by Kral (1927) and Pful (1866) were prepared in a digital version and published by the Sorbian Institute in 2006.[8] Unfortunately, this project has not been developed further technically, and some users may experience difficulties while trying to reach the server.

The small digital US-Czech dictionary (Martínek & Brankačkec, 2005) could also be helpful for our purposes. It was created by manually choosing about 10,000 entries from the US spelling dictionary by Völkel (1981) and translating the German part into Czech. A modest US-Czech dictionary (but also a LS-Czech one and even a Polabian-Czech glossary) is also available at D. Krčmařík's personal website,[9] unfortunately without any more detailed information.

A crowdsourced multilingual dictionary Glosbe[10] also includes US. It is not surprising that the largest amount of data can be found in the US-German (6,674 phrases; 1,133 examples) and US-English (6,454 phrases; 231 examples) sections. The US-Czech part, considering its limited list of entries (4,863 phrases; 252 examples – cf. US-Polish: 5,414/233), may be inspirational principally in terms of web design and engaging the lay public into the project; as regards the data itself, it is deficient in too many ways.

Another crowdsourced project – Wiktionary – also has its US version[11] (the LS version doesn't exist yet), with 4,176 entries.

---

kucik/index.html

[5] https://www.serbski-institut.de/os/Geografiske-mjena-hornjoserbsce/

[6] https://soblex.de/sotra/

[7] https://hornjoserbsce.de/dow/

[8] http://www.serbski-institut.de:8180/dict/online

[9] http://slovnik.vancl.eu/abc/index.php

[10] https://hsb.wiktionary.org/wiki/H%C5%82owna_strona

[11] https://app.glosbe.com (All statistics from the portal were valid as of 8 April 2021.)

A few remarks about the metalexicographic aspect: New dictionaries elicit only a little attention among language professionals, and there is a rather languorous discussion about these developments. Most of the dictionaries are reviewed only once, e.g., the English-US (Wornar, 2007) and the US-English (Stone, 2005) dictionaries were only examined by Szpila (2008), albeit in "unusual" detail. The German-US dictionary was reviewed twice (Lewaszkiewicz, 2008; Šěrakowa, 2009). A more vivid discussion can be observed on the so-called "new Völkel" (2005; see Pohončowa & Šołćina, 2006 for its review) that reflects changes in the US orthography, especially of more recent loanwords from German and English and with a change even in the alphabetic order: the grapheme *ć* was formerly arranged after *t*; newly, it follows *č*.

1.2 Jiří Mudra's US-Czech dictionary

Jiří Mudra (1921–2009) was a doyen of Sorbian studies in former Czechoslovakia. Besides undertaking numerous activities in organising, propagating, and interpreting Lusatian literature and other cultural artifacts (Kaleta, 2011), he was active in Sorbian linguistics. He is a co-author of a four-volume US textbook (Mudra & Petr, 1982–1989), and for many decades he also worked on a US-Czech dictionary, which unfortunately remained unfinished. He left behind a manuscript of approximately 22,000 entries ranging from *a* to *smyknyć*, plus its digitised version. In his paper, regrettably too general and brief, Mudra himself characterised his planned *chef d'ouvre* (Mudra, 1999). He intended to make a dictionary that would be helpful for all Czech speakers interested in US. It should have been medium-sized, with approximately 30–40 thousand entries covering all grammar words and the most frequent content words in US; the frequency criterion could not be taken into account reliably due to the lack of empirical data. Although there were some US-Czech dictionaries, they were outdated, and their lists of entries were very limited along with having a primitive microstructure (Páta, 1920; Mohelský, 1948), an thus they would suffice to only cover users' elementary needs. Presumably, this was the primary impetus for Mudra to start work on his dictionary. Besides its own excerptions, it was based on then-recent US dictionaries (Völkel, 1981; Budarjowa, 1990; Korjeńk, 1995; Mehrowa & Pawlikowa, 1995; Trofymovyč, 1974; Jentsch et al., 1989, 1991) as well as some older ones (Jakubaš, 1954; Kral, 1927; even Pful, 1866), along with US grammar books (Faßke & Michalk, 1981; Šewc-Schuster, 1976, 1984). However, no corpora data were available to Mudra, unlike the current situation, which gives us a chance to finish the project.

2. Completing Mudra's dictionary

In this chapter, we present our plan to complete the main body of Mudra's dictionary. Naturally, the suggestions described below are mere theses – partly due to the limited space of the paper, and partly because we do not present a ready-made style guide for the lexicographic team (this should be created later, but we are still in the initial phase of the project). We will successively deal with various aspects: the current state of the

project (2.1) and its possible pitfalls (2.6), technical issues (2.2), other potential data sources (2.3), specific proposals for changes in both the macrostructure and microstructure of the dictionary (2.4) as well as purely practical issues, such as the workflow of the whole project (2.5). We constantly consider the target users (especially 2.1); no matter how we define them – until the resulting dictionary will have been made available to the users, Mudra's work is not complete, thus *de facto* worthless.

2.1 Mudra's dictionary from our perspective

Currently, Mudra's lifelong lexicographic work looks like this:

Figure 1: Jiří Mudra's lexicographical legacy

Such a view may arouse nostalgia for old-school lexicography, albeit along with regret about the incompleteness of the project. It does not matter how much data Mudra has collected and how good this data is. The work done to this point is practically useless until the dictionary starts to serve its users: specifically, if it is the only dictionary of its kind or if no decent dictionary exists in this language combination. Of course, the question arises whether it might not be better to start again from scratch, without the "burden" of pre-corpus lexicography. However, we do not want to do this – for several

reasons. Firstly – and above all – we do not have sufficient human resources for a completely new project. Also, the US is still a low-resource language, with no representative sample of the current language of adequate quality available (see the Hotko corpus in 2.3). Furthermore, Mudra's material is too valuable to ignore, although not corpus-based. In the vast majority of it, it adequately represents the desired *equivalence* between the US and Czech lexicons. Any inaccuracies and errors can simply be revised; this task is definitely easier than pointless – not harmless! – drudgery. Last but not least, we consider Mudra an indisputable personality of Sorbian studies in Czechia, who deserves respect and credit. These are the main reasons why we prefer completing Mudra's dictionary, considering ourselves as editors of his preprocessed material, successors to his unfinished work. It does not mean that the work will proceed without a professional lexicographic and critical approach. The assembled data needs to be revised and supplemented with new knowledge from new sources (at least those from the last twenty years) which were not available to Mudra or deliberately excluded by him. In addition, it is necessary to process the rest of the alphabet (the rest of the letter S, followed by the letters Š–Ž).[12] Eventually, the outcome needs to be passed on in an acceptable form to users as soon as possible. We already consider the paper dictionary to be an obsolete form, preferring an electronic one, which, among others, allows the publishing of new entries progressively as they emerge or to update already published entries easily.

We define the target users broadly enough due to the specificity of the language combination. In such a situation, a universally designed dictionary accumulates the functions of different types of dictionaries (terminological, phraseological, vernacular, etc., but often conversation books too; cf. Mudra, 1999: 260). Therefore, we primarily aim at any Czech-speaking person interested in US (enthusiasts, linguists and other humanities scholars, translators, etc.) as well as Sorbian people interested in Czech. The dictionary concept should correspond to this: it should be universal enough to serve the practical needs of US students, along with those of people traveling to Lusatia, translators of US fiction, etc. Besides, the project's openness to the general public (2.5) explicitly counts on its active participation and can respond flexibly to its needs.

2.2 Technical issues

The CD in the lower right part of the photograph above suggests that the data has been digitised. Indeed, we have an MS Word file with the semi-finished dictionary,

---

[12] As Völkel suggests (approximately 52,000 entries on 661 pages), the remaining part covers almost *one-third* of the whole dictionary. Hopefully, this is sufficient proof we intend not just to publish the main body of Mudra's dictionary without making any contribution, but we understand our engagement as an equal partnership with a deceased colleague.

which looks like this:

CH

chabłak 2m *váhavec, vrtkavý člověk*
chabłanje 18s *kolísání*; ~ temperatury, deklinacije cuzych słowow ~ *teploty, skloňování cizích slov*
chablar 4m *váhavec, vrtkavý člověk*
chabłatosć 13ž *rozkolísanost*
chabłaty 21 1 *vrtkavý, vratký* 2 *nerozhodný*
chabła¦ć 39 ned 1 *kymáce¦t se*; štomy -ja we wětřiku *stromy se -jí ve větru*; -ca strowota *kolísavé zdraví* 2 *váhat, otálet*; w swojim rozsudźe ~ *otálet s rozhodnutím*
chabław¦y 21 *nerozhodn¦ý*; -a powaha *-á povaha*
chablo¦tać 39 j1 -tam /-cem ned = chabłać
chabliwy 21 = chabławy|
chagrin 1m odb *šagrén*
chacho¦tać 39 j1 -tam/-cem ned *chechtat se*
chalupa 7ž kniž řidč *chalupa*
champion 1m *šampion, přeborník*

Figure 2: An example from Mudra's unfinished dictionary in digitised form (MS Word), cf. Mudra, 1999: 261–262

As can be seen in the figure, the microstructure of the entry is quite simple.[13] The lemma is followed by grammatical information (numeric or alphameric code assigns the lexeme to the appropriate class of words, irregular and problematic forms appear) and a stylistic marker. The semantic part follows, i.e., an overview of meanings – or their Czech equivalents – supplemented sometimes by a few significant collocations or idioms. Figure 2 is an example of a typical lexicographic production in the Czech environment in the second part of the 20th century: an apparent effort to save space as much as possible is made due to the limited space of the printed dictionary. Nevertheless, we no longer have these limits today. Therefore, we can enrich entries with missing features (see 2.4 below), add more collocations, replace the opaque system of ciphers and codes with explicit metalanguage, etc.

We feel that continuing to work in MS Word or another text editor is an anachronism now, especially if numerous dictionary writing systems (DWS) are available, which have become the current lexicographic standard. The final choice has not yet been made, although some team members have good experience with the DWS TshwaneLex while working on another dictionary (Škrabal, 2016).

---

[13] Even simpler is the microstructure of entries in Mudra & Petr (1989), which is understandable, for it is not a full-fledged dictionary, just a practical tool for Czech students of US (it covers approximately 13,000 entries used in Mudra and Petr's textbook of US). Mudra intended it as a predecessor to a regular dictionary (Mudra, 1999: 260).

We primarily propose an online dictionary, and its website[14] should be responsive so that it can be used on various devices, not just a desktop computer. In the future, a stand-alone application for smartphones should be created too. A printed version is not foreseen, although processing the data does not necessarily exclude it. However, it will be a minor mode of using the data in a print-on-demand scenario if a potential applicant makes an explicit request. Instead, we imagine using printed materials for didactic purposes: the teacher prints out only a pertinent fragment from the dictionary for students, such as topic-related vocabulary for a relevant lesson, or lexemes belonging to the same word family, etc. This can be handled simply via tick-box features in most DWS and the subsequent filtering out of the selected entries.

2.3 Other data sources

Mudra's materials are in themselves partly a compilation (in the best sense of the word) of various dictionaries (1.1–1.2), supplemented by long-term excerpts, but these are inherently selective. Therefore, they should be challenged by other sources available today, mainly corpus data, and ideally, that of parallel corpora. Currently, the Hotko v2 corpus of US is available (via the KonText interface, see 2.5 below). It contains journalistic (57%), fiction (23%), religious, and scientific texts from the middle of the 19th century to the present that were scanned and OCR-ed, but not corrected. Besides, the corpus is neither lemmatised nor morphologically annotated, which can complicate the search. Some of the data, e.g., a number of dictionaries (12%), old texts reflecting historical spelling, or numerous German-language fragments, are unusable for our purpose. The total corpus size is currently 43.9 million tokens, including punctuation; more than half of the texts (54%) date from the 1990s onwards. Post-war texts seem relevant to us (with exceptions mentioned above), and such a subcorpus contains no more than 31 million tokens.

Later, it will also be possible to start using data from the parallel US-Czech corpus, which is planned as an extension of the InterCorp (IC) corpus (Čermák & Rosen, 2012). We currently have the first three texts ready for alignment (one of which is a translation from US, and two into US). We believe that IC is a good investment, as not only the US-Czech subcorpus is being created, but – with a suitable selection of texts[15] – also the subcorpora of other US-X language combinations. These can then serve as a dataset for other translation dictionaries.

---

[14] We are counting here on our own website although, in theory and after mutual agreement, it would be possible to connect to the already existing infrastructure, such as the above-mentioned Soblex dictionary, that nowadays is the US-German-US dictionary only. The portal could be extended by several language modules (Czech, Polish, …), with US as a pivot language.

[15] The ideal dataset would primarily consist of as many US originals as possible, but these are only seldom translated into foreign languages. For example, Jurij Brězan's novel *Stary nan* can appear in IC in the Czech, Russian, or Slovak translations, while the US translation of Douglas Adams' *The Hitchhiker's Guide to the Galaxy* could be aligned to another 25 languages along with the English original.

2.4 Amendments in the macrostructure and microstructure of Mudra's dictionary

The basic data structure (Fig. 2) will be converted to DWS, and other desirable elements and attributes will extend the DTD structure. It is necessary, among other tasks, to predefine various grammatical and semantic classes of words, etc. Frequencies should play an important role, as they were not taken into account in pre-corpus times, at least in the Czech environment.[16] After all, Mudra himself was aware of the importance of frequency data; in his paper (1999: 261) he expressed regret about the non-existence of both a US frequency dictionary and a dictionary of spoken US. He also could not rely on corpus data in his work, as the first US publicly available corpora did not appear until 2013 (Hotko v1). Nevertheless, we believe that he would appreciate the corpus as an amazing tool for lexicographers.

We do not want to overestimate the frequency data, being well aware of the problems of corpus data (2.3) that should thus be treated with caution. However, it is possible to cover frequency information in several ways:

- by dividing the lexicon into a reasonable number of frequency bands (e.g., *very common – common – unusual – sparse/idiosyncratic*);

- by extracting the top frequency lexemes (e.g., the top 1,000 words) reliably even on a relatively small corpus, such as Hotko, and then marking them directly in the dictionary with a proper graphic means (suitable mainly for didactic purposes);

- by stating the frequency data (i.p.m. preferably) explicitly for selected words.

We want to extend the number of collocations, especially idioms; the exemplification should be the central part of the revisited entry. Appropriately selected examples from IC will be quoted along with their Czech translations; examples from non-parallel corpora will be provided with ad-hoc translations. Besides, the user can find more exemplification directly in a given corpus via a hypertext link.

The pronunciation is entirely neglected by Mudra as he considers it to be regular, perhaps with some rare exceptions and in the case of loanwords. We see a wise solution in audio recordings directly from native speakers, whether professionals or amateurs.[17] This is the area where we believe crowdsourcing can be applied most effectively (see also 2.5 below).

The pragmatic aspect of word usage is also essential: our dictionary should inform the users about the word's specific place within the lexicon and even warn them of possible

---

[16] With an exception of a few markers denoting either the uncommon or obsolete usage of words.

[17] Cf. the US section of the pronunciation guide Forvo.com: https://forvo.com/languages/hsb/

negative reactions (e.g., to offensive and vulgar words). Usage notes are not a common phenomenon in the Czech lexicographic tradition (Šemelík & Škrabal, 2019), and we understand their inclusion in our dictionary as partial repayment of this debt. In a specific US-Czech combination, notes can also be used to alert false friends explicitly or, in general, any potentially problematic places for a Czech user.

As far as the macrostructure of the dictionary is concerned, it will also undergo some revisions and add-ons. Mudra avoided some (from his point of view) problematic groups of words in the list of entries, such as Germanisms or vulgarisms, tending towards a literary language that may sound somewhat artificial today. We cannot identify with this protective approach, as it does not correspond to our descriptive basis. The lexicographic description should be in accordance with the language reality. Any language taboo is also inadmissible in our eyes: vulgar words have a valid place in the lexicon of each language and should be therefore described, although with special treatment and means (be it markers, usage notes, or other features).

Above all, the current colloquial form of US should be bolstered, not withdrawing from new loanwords from both German and English. Naturally, data from the US spoken corpora would be precious, yet, this is still a long way off, and we must be content even with data from written corpora and the most recent US dictionaries such as Völkel (2005) and Jenč et al. (2006).

2.5 Workflow

Our project's workflow depends mainly on the amount of funding available within the appropriate grant. We expect the involvement of two academic workplaces (Institute of Slavonic Studies of the Czech Academy of Sciences and Institute of the Czech National Corpus) and one civic body (Society of Friends of Lusatia). The former will provide the relevant know-how and infrastructure [18] while engaging the lay public is an opportunity to use the "wisdom of crowds" (Surowiecki, 2005). Amateur volunteers can contribute to us in various ways: be it by notifying us of errors in the dictionary, suggesting changes or additions in the list of entries or individual entries, recording the pronunciation, etc. We also want to involve Czech students of Sorbian studies[19] (or Sorbian students of Czech); they could learn the basics of practical lexicography and

---

[18] ISS has numerous and long-term experience in the field of Slavic lexicography. The Russian-Czech electronic dictionary database with the Large Czech-Russian Dictionary (http://slovnik.slu.cas.cz), as well as the digital portal in Old Slavonic *Gorazd* (http://gorazd.org/gulliver/), are digitally accessible to the public. Besides its major project (Czech National Corpus, aimed at studying the Czech language), ICNC provides infrastructure for parallel corpora, including the emerging US component (see 2.3 above), and it also hosts the Hotko corpus via the KonText interface (Machálek, 2014).

[19] Currently implemented at the Faculty of Arts, Charles University in the form of an optional practical course in US. In the first year, it was attended by approximately 10 students; in distance learning during the pandemic, the number of students decreased to 3–5.

related linguistic disciplines (lexicology, corpus linguistics, etc.) during a suitably designed course or a hands-on workshop.

2.6 Problematic issues

In addition to the already mentioned need for grant support, the most severe difficulty concerns the copyright to Mudra's work. Negotiations with the heirs of these rights are not without problems. The idea of the dictionary as a tangible book artifact is too entrenched in laypeople, and its virtual form is difficult to accept. (Although more and more people do not even use paper dictionaries, which are too cumbersome for them, and prefer online resources.) The fear of insufficient acknowledgment of Mudra's life-long effort, of the appropriation of his work and merits probably also plays a role. However, these fears seem odd: all of Mudra's data used by us will be appropriately marked in the dictionary (by an icon, cipher, etc.), and his name will, of course, be listed *first* in the list of used sources. Besides, we want to place Mudra's biographical profile on the dictionary's website, which would summarise his life and Sorabist career (1.2).[20]

Despite certain complications, we believe in the successful outcome of the negotiations. Without this, Mudra's work – undoubtedly remarkable and worthy of respect – will remain a mere fragment (Fig. 1 and 2) instead of fulfilling its purpose and receiving the attention and recognition it deserves.

3. Conclusion

To answer the question in the title of this paper: what can be done about Mudra's lexicographic "posthumous child"? A lot – under two basic assumptions: a) that we obtain the permission of the copyright heirs for the free (yet not arbitrary) handling of data, and b) that we obtain sufficient grant support. In such a case, we can complete the unfinished work, which otherwise would end up irretrievable and useless, and thus give Mudra's project appropriate credit in the Czech Sorabist milieu, even among the youngest generation which no longer remembers or knew the doyen. It is evident that the material assembled by Mudra (as by anyone else) needs to be approached critically, with hands not bound by exaggerated piety. The aim is to reconcile this material with both the additional data sources and the modern lexicographic approach, as we have tried to describe in this paper. Only in this way can the resulting dictionary begin to fulfil its primary purpose, i.e., to serve its users as effectively as possible. There may not be a vast number of them (we estimate it as in the hundreds to low thousands), but this does not reduce their need for quality lexicographic tools.

---

[20] Regrettably, Jiří Mudra does not have his own article on either the Czech or US Wikipedia yet; the most complete biographical source is an anthology, published posthumously in homage to Mudra (Kaleta, 2011).

4. Acknowledgements

5. References

Bartels, H. & Kaulfürst, F. & Szczepański, M. & Wölke, S. (2021). Das Monitoring des sorbischen Schrifttums. Grundlagen und erster Jahresbericht eines neuen Forschungsvorhabens. *Lětopis* 68(1), pp. 77–136.

Bayer, M. (2006). *Sprachkontakt deutsch-slavisch. Eine kontrastive Interferenzstudie am Beispiel des Ober- und Niedersorbischen, Kärntnerslovenischen und Burgenlandkroatischen.* Frankfurt am Main: Peter Lang.

Bejmak W. & Schmiedelowa, A. & Szczepański, M. & Wölkowa, S. (forthcoming). *Němsko-hornjoserbski słownik na Hornjoserbsce.de.* Budyšin: Serbski institut, Accessed at: https://obersorbisch.de/dow/info (22 March 2021).

Budarjowa, L. (1990). *Wörterbuch Obersorbisch-deutsch / Słownik hornjoserbsko-němski.* Bautzen: Domowina-Verlag.

Čermák, F. & Rosen, A. (2012). The case of InterCorp, a multilingual parallel corpus. *International Journal of Corpus Linguistics*, 17(3), pp. 411–427.

Faßke, H. & Michalk, S. (1981). *Grammatik der obersorbischen Schriftsprache der Gegenwart. Morphologie.* Bautzen: Domowina-Verlag.

Hajduk-Veljkovićowa, L. (2017). *Kak to jenož praju? Přiručka za bohatši słowoskład za wyšu šulu a gymnazij.* Budyšin: Domowina.

Hornjoserbsce.de/obersorbisch.de. Budyšin: Serbski institut. Accessed at: https://obersorbisch.de/dow/info (22 March 2021).

Itoya, B. (2013). Badania metod kodyfikacji leksyki górnołużyckiej w zestawieniu z dorobkiem metaleksykografii i leksykografii czeskiej. In M. Milewska-Stawiany & S. Wölkowa (eds.) *Leksikologiske přinoški II. IV. Seminar serbskeje słowotwórby a leksiki. IV Seminarium Słowotwórstwa i Słownictwa Łużyckiego. Uniwersytet Gdański / Serbski institut 31. 5. – 1. 6. 2012.* Budyšin: Serbski institut, pp. 102–115.

Jakubaš, F. (1954). *Hornjoserbsko-němski słownik. Obersorbisch-deutsches Wörterbuch.* Budyšin: Domowina.

Jenč, H. & Pohončowa, A. & Šołćina, J. (2006). *Němsko-hornjoserbski słownik noweje leksiki. Deutsch-obersorbisches Wörterbuch neuer Lexik.* Bautzen: Domowina-Verlag.

Jentsch, H. & Michalk, S. & Schierack, I. (1989, 1991). *Deutsch-Obersorbisches Wörterbuch A–K/ L–Z.* Bautzen: Domowina-Verlag.

Kaleta, P. (2011). *Jiří Mudra. K 90. výročí narození českého vlastence a přítele Lužických Srbů (s albem fotografií).* Praha / Budyšin: Společnost přátel Lužice / Maćica serbska.

Korjeńk, J. (1995). *Terminologija za předmjet fyzika: němsko-hornjoserbsce, hornjosebsko-němsce.* Budyšin: Domowina.

Kral, J. (1927). *Serbsko-němski słownik hornjołužiskeje serbskeje rěče. Sorbisch-*

deutsches Wörterbuch der Oberlausitzer sorbischen Sprache. Budyšin: Maćica serbska.

Krčmařík, D. *HLS slovník.* Accessed at: http://slovnik.vancl.eu/abc/index.php (22 March 2021).

Leszcyński, R. (2013). *Podręczny słownik polsko-dolnołużycki/dolnołużycko-polski. Pśirucny słownik pólsko-dolnoserbski/dolnoserbsko-pólski.* Budyšin: Domowina.

Lewaszkiewicz, T. (2008). Recenzje. Helmut Jentsch / H. Jenč, Anja Pohontsch / A. Pohončowa, Jana Schulz / J. Šołćina, Deutsch-obersorbisches Wörterbuch neuer Lexik/ Němsko-hornjoserbski słownik noweje leksiki. Domowina-Verlag / Ludowe nakładnistwo Domowina, Bautzen / Budyšin 2006. *Slavia Occidentalis* 65, pp. 163–164.

Machálek, T. (2014). *Kontext – rozhraní pro vyhledávání v korpusech.* Praha: Filozofická fakulta Univerzity Karlovy. https://kontext.korpus.cz (8 April 2021).

Martínek, F. & Brankačkec, K. (2005). *Hornjoserbsko-čěski słownik.* Accessed at: http://www.serbski-institut.de:8180/dict/ and: https://www.serbski-institut.de/de/Online-Publikationen/ (22 March 2021).

Mehrowa, H. & Pawlikowa, B. (1996). *Terminologija za předmjet geografia: němsko-hornjoserbsce, hornjosebsko-němsce.* Budyšin: Domowina.

Meškank, T. (2017). *Serbske předmjena. Serbske pśedmjenja. Sorbische/wendische Vornamen.* Bautzen: Witaj-Sprachzentrum.

Mohelský, V. (1948). *Mluvnice hornolužické srbštiny a slovník hornosrbsko - český = Rěčnica hornjołužiskeje serbśćiny a słownik hornjoserbsko-čěski.* Olomouc: V. Tomek.

Mudra, J. & Petr, J. (1982, 1987, 1989). *Učebnice lužické srbštiny.* Praha: Filozofická fakulta Univerzity Karlovy.

Mudra, J. (1999). K problematice hornjoserbskeho-čěskeho słownika. *Lětopis* 46. *Wosebity zešiwk*, pp. 260–263.

Páta, J. (1920). *Kapesní slovník lužicko-česko-jihoslovanský a česko-lužický.* Praha: Českolužický spolek „Adolf Černý".

Pful, K. B. (1866). *Łužiski serbski słownik.* Budyšin: Maćica serbska.

Pohončowa, A. (2008). Neologizmy w hornjoserbskej spisownej rěči a jich leksikografiske fiksěrowanje. In: M. Milewska-Stawiany & S. Wölkowa (eds.) *Leksikologiske přinoški: III. Seminar serbskeje słowotwórby.* Budyšin: Domowina, pp. 43–49.

Pohončowa, A. & Šołćina, J. (2007). Strategije modernizowanja hornjoserbskeho słowoskłada. *Rozhlad* 57(1), pp. 5–9.

Pohončowa, A. & Šołćina, J. (2006). Nowy prawopisny słownik – wšitko jasne? (= Das neue Orthographiewörterbuch – alles klar?). *Serbska šula* 59(1), pp. 11–12.

Schuster-Šewc, H. (1976, 1984). *Gramatika hornjoserbskeje rěče I, II.* Budyšin: Domowina.

Serbski Institut Budyšin (2021). *HOTKO: hornolužický textový korpus, verze 2 z 6. 3. 2021.* Praha: Ústav Českého národního korpusu. Accessed at: https://www.korpus.cz (22 March 2021).

Serbski Institut Budyšin (2021). *SobLex. Hornjoserbsko-němski słownik. V3.03.06*

14.01.2021. Accessed at: https://www.soblex.de (22 March 2021).

Siatkowska, E. & Leszczyński, R. (2002). *Słownik polsko-górnołużycki i górnołużycko-polski. Pólsko-hornjoserbski a hornjoserbsko-pólski słownik.* Warszawa: Uniwersytet Warszawski, Instytut Filologii Słowiańskie.

Surowiecki, J. (2005). *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations.* New York: Doubleday.

Szpila, G. (2008). Edward Wornar: Jendźelsko-hornjoserbski šulski słownik. English-Upper Sorbian Learner's Dictionary. Recensija. *Lětopis* 55(1), pp. 158–163.

Szpila, G. (2014). Rafał Leszczyński: Podręczny słownik polsko-dolnołużycki/dolnołużycko-polski. Pśirucny słownik pólsko-dolnoserbski/dolnoserbsko-pólski. *Lětopis* 61(2), pp. 157–162.

Šemelík, M. & Škrabal, M. (2019). Poznámky k poznámkám. Usage notes v českém lexikografickém prostředí. *Naše řeč* 102(1), pp. 25–35.

Šěrakowa, I. (2009). Moderne hornjoserbske słownistwo we wužitnym słowniku. *Rozhlad* 59(6), pp. 222–224.

Škrabal, M. (2016). Straddling the Boundaries of Traditional and Corpus-Based Lexicography: A Latvian-Czech Dictionary. In T. Margalitadze & G. Meladze (eds.) *Proceedings of the XVII EURALEX International Congress. Lexicography and Linguistic Diversity.* Tbilisi: Ivane Javakhishvili Tbilisi State University, pp. 910–914.

Šołćina, J. (1999). Recenzija: Rafał Leszczyński: Hornjoserbsko-pólski a pólsko-hornjoserbski słownik pozdatnych ekwiwalentow / Górnołuzycko-polski i polsko-górnołuzycki słownik ekwiwalentów pozornych, Warszawa 1996. *Lětopis* 46(1), pp. 157–161.

Trofymovyč, K. K. (1974). *Hornjoserbsko-ruski słownik = Verchnelužicko-russkij slovar'.* Budyšin / Moskva: Domowina / Russkij jazyk.

Völkel, P. (1970–1981). *Prawopisny słownik. Hornjoserbsko-němski słownik. Obersorbisch-deutsches Wörterbuch.* Budyšin: Domowina.

Völkel, P. (2005). *Prawopisny słownik. Hornjoserbsko-němski słownik. Obersorbisch-deutsches Wörterbuch.* Budyšin: Domowina.

Wölkowa, S. & Pohončowa, A. & Šołćina, J. Accessed at: https://hornjoserbsce.de/kuciki/ (22 March 2021).

Wornar, E. (2007). *Jendźelsko-hornjoserbski šulski słownik. English-Upper Sorbian Learner's Dictionary.* Budyšin: Domowina.