Univerzita Palackého v Olomouci
Přírodovědecká fakulta
Katedra geoinformatiky

PROSTOROVÁ PODMÍNĚNOST VYBRANÝCH PŘÍJMENÍ A NÁŘEČNÍCH VARIANT NA ÚZEMÍ ČESKA

Diplomová práce

Bc. Vojtěch BARTOŠ

Vedoucí práce
prof. RNDr. Vít Voženílek, CSc.

Olomouc 2024
Geoinformatika a kartografie

ANOTACE

Předkládaná diplomová práce zkoumá prostorovou podmíněnost vybraných příjmení a nářečních variant v České republice s využitím moderních geoinformačních systémů (GIS) a kombinací metod prostorové a statistické analýzy. Cílem práce je stanovit, zda existuje souvislost mezi geografickým rozmístěním specifických příjmení a nářečními oblastmi České republiky. V rámci práce byly vytvořeny nové metriky a nástroje pro analýzu a vyhodnocení jazykovězeměpisných dat. Teoretická část je zaměřena na studium metod prostorové analýzy a jejich aplikace v onomastice a dialektologii. Jako vstupní data byla využita data o rozmístění vybraných příjmení z webové aplikace KdeJsme.cz a z vektorizovaných map vybraných kapitol Českého jazykového atlasu. Pomocí vytvořených metrik byla na základě analýzy těchto dat vyhodnocena prostorová podmíněnost vybraných příjmení a nářečních variant. Práce přináší nový pohled na zkoumání vztahů mezi příjmením a dialektologickými daty. Výsledky práce mají potenciál pro uplatnění v dalších výzkumech v oblastech dialektologie, demografie a jazykové geografie. Diplomová práce také představuje novou metodiku, která může být aplikována v dalším výzkumu.

KLÍČOVÁ SLOVA

prostorová podmíněnost; příjmení; nářečí; jazykový zeměpis; dialektologie

Počet stran práce: 69

Počet příloh: 5 (z toho 4 volné a 1 elektronická)

ANOTATION

This thesis examines the spatial conditionality of selected surnames and dialectal variants in the Czech Republic using modern geographic information systems (GIS) and a combination of spatial and statistical analysis methods. The objective of the thesis is to determine whether there is a connection between the geographical distribution of certain surnames and the dialectal areas in the Czech Republic. For the purpose of this thesis new metrics and tools have been developed to analyse and evaluate geolinguistic data. The theoretical part focuses on the study of spatial analysis methods and their application in onomastics and dialectology. The input data included the distribution of selected surnames from the web application KdeJsme.cz and from vectorised maps of selected chapters of the Czech Language Atlas (Český jazykový atlas). Based on the analysis of these data, the developed metrics were used to evaluate the spatial conditionality of selected surnames and dialect variants. The thesis provides a new perspective on the study of relationships between surnames and dialectological data. The results of the research conducted for the purpose of this thesis offer promising avenues for further research in the field of dialectology, demography or linguistic geography. The thesis also introduces a new methodology that can be used in further research.

KEYWORDS

spatial conditionality; surnames; dialects; geolinguistics; dialectology

Number of pages 69
Number of appendixes 5

Prohlašuji, že

- diplomovou práci včetně příloh, jsem vypracoval samostatně a uvedl jsem všechny použité podklady a literaturu.
- jsem si vědom, že na moji diplomovou práci se plně vztahuje zákon č.121/2000 Sb. - autorský zákon, zejména § 35 – využití díla v rámci občanských a náboženských obřadů, v rámci školních představení a využití díla školního a § 60 – školní dílo,
- beru na vědomí, že Univerzita Palackého v Olomouci (dále UP Olomouc) má právo nevýdělečně, ke své vnitřní potřebě, diplomovou práci užívat (§ 35 odst. 3),
- souhlasím, že údaje o mé diplomové práci budou zveřejněny ve Studijním informačním systému UP,
- v případě zájmu UP Olomouc uzavřu licenční smlouvu s oprávněním užít výsledky a výstupy mé diplomové práce v rozsahu § 12 odst. 4 autorského zákona,
- použít výsledky a výstupy mé diplomové práce nebo poskytnout licenci k jejímu využití mohu jen se souhlasem UP Olomouc, která je oprávněna v takovém případě ode mne požadovat přiměřený příspěvek na úhradu nákladů, které byly UP Olomouc na vytvoření díla vynaloženy (až do jejich skutečné výše).

V Olomouci dne 8. května 2024 jméno autora: Vojtěch Bartoš

podpis: 

Děkuji vedoucímu práce prof. **RNDr. Vítu Voženílkovi, CSc.** za přínosné podněty a připomínky k vypracování práce, stejně jako děkuji za ochotu přizpůsobit konzultace mým časovým i geografickým možnostem.

Dále děkuji odborné konzultantce práce **PhDr. Martině Ireinové, Ph.D.** z Ústavu pro jazyk český Akademie věd České republiky za cenné rady, trpělivost, a hlavně za poskytnutý čas, který mi věnovala při odborných konzultací.

V neposlední řadě děkuji své rodině a blízkým za veškerou podporu, kterou mi poskytovali nejen během vypracování této práce, ale také v průběhu celého studia.

Diplomová práce byla vypracována v rámci projektu NAKI II "Atlas českého jazyka 2027: celoúzemní výzkum nářečí českého jazyka po 50 letech", reg. č. DH23P03OVV050.

Zadání – 1. strana

UNIVERZITA PALACKÉHO V OLOMOUCI

Přírodovědecká fakulta

Akademický rok: 2022/2023

ZADÁNÍ DIPLOMOVÉ PRÁCE

(projektu, uměleckého díla, uměleckého výkonu)

Jméno a příjmení:	Bc. Vojtěch BARTOŠ
Osobní číslo:	R220651
Studijní program:	N0532A330009 Geoinformatika a kartografie
Téma práce:	Prostorová podmíněnost vybraných příjmení a nářečních variant na území Česka
Zadávající katedra:	Katedra geoinformatiky

Zásady pro vypracování

Cílem diplomové práce je zjistit prostorovou podmíněnost mezi vybranými příjmeními a jejich nářečními variantami na území Česka. Student připraví k prostorovým analýzám dvě sady dat. První budou tvořit vrstvy rozmístění osob s vybranými příjmeními, např. Vrabec x Brabec nebo Odehnal x Vodehnal, druhou sadu dat pořídí digitalizací z map Českého jazykového atlasu příslušných nářečních variant slov tvořících příjmení. Prostorovými analýzami potvrdí či vyvrátí domněnku, že bydliště osob se specifickými příjmeními mají vazbu na nářečí oblasti českého jazyka. Student bude spolupracovat s dialektology ÚJČ AV ČR a případně zahrne do analýz i příbuzná témata, např. rodáky.

Rozsah pracovní zprávy:	max. 50 stran
Rozsah grafických prací:	dle potřeby
Forma zpracování diplomové práce:	elektronická

Seznam doporučené literatury:

BALHAR, Jan a JANČÁK, Pavel. Český jazykový atlas [online]. 2, elektronické, opravené, doplněné vyd. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, 2021. ISBN 978-80-86496-66-5. Dostupné z: https://cja.ujc.cas.cz.

MALAČKA, Ondřej. KdeJsme.cz. [online]. 2011. Dostupné z: https://www.kdejsme.cz/.

MATEOS, Pablo a TUCKER, Ken. Forenames and Surnames in Spain in 2004. [online]. 2008. *Names*. 2008, 56, 3, s.165–184. Dostupné z: https://doi.org/10.1179/175622708X332860.

UPTON, Clive a WIDDOWSON, J.D.A. An Atlas of English Dialects. 2. vyd. Abingdon: Routledge. 2006. ISBN 978-0-415-39233-4.

VOŽENÍLEK, Vít a KAŇOK, Jaromír a kol. Metody tematické kartografie: vizualizace prostorových jevů. 1. vyd. Olomouc: Univerzita Palackého v Olomouci pro katedru geoinformatiky, 2011. ISBN 978-80-244-2790-4.

Vedoucí diplomové práce: **prof. RNDr. Vít Voženílek, CSc.**
Katedra geoinformatiky

Zadání 2. strana

Datum zadání diplomové práce: **5. prosince 2022**
Termín odevzdání diplomové práce: **9. května 2024**

L.S.

doc. RNDr. Martin Kubala, Ph.D.
děkan

prof. RNDr. Vilém Pechanec, Ph.D.
vedoucí katedry

V Olomouci dne 1. září 2023

OBSAH

SEZNAM POUŽITÝCH ZKRATEK

Zkratka	Význam
AVČR	Akademie věd České republiky
ČJA	Český jazykový atlas
GIS	geografický informační systém
MVČR	Ministerstvo vnitra České republiky
ORP	obec s rozšířenou působností
ÚJČ	Ústav pro jazyk český

ÚVOD

Jména a příjmení, která nosíme, slouží nejen jako základní stavební kameny naší identity, ale jsou rovněž okny do kultury a prostředí, v níž se vyvíjela. Česká republika, se svou rozmanitostí nářečí a jazykových jevů, poskytuje vhodné prostředí pro zkoumání vztahu mezi jazykem, příjmením a prostorem, ve kterém jsou rozmístěny. Tato diplomová práce se věnuje analýze prostorové podmíněnosti vybraných příjmení a specifických nářečních variant na území České republiky a jejím cílem je určit, zda lze prokázat spojitost mezi rozmístěním jednotlivých příjmení a nářečními oblastmi.

Výzkum se zaměřuje na identifikaci vzorců v rozmístění příjmení, na zpracování nářečních dat a následnou analýzu těchto dvou datových sad za účelem odhalení možných korelací. K tomuto účelu byly využity moderní geoinformační technologie (GIS) a kombinace metod prostorové a statistické analýzy, což umožňuje detailní vizualizaci a kvantitativní hodnocení zkoumaných fenoménů. Tento přístup umožňuje lépe pochopit, jak nářečí a používaný jazyk ovlivňovaly a stále ovlivňují a rozložení a dynamiku příjmení v různých částech země.

V rámci práce byly vytvořeny dvě nové metriky pro vyhodnocování jazykovězeměpisných dat. Tyto metriky a nástroje pro její implementaci do výzkumu jsou v práci detailně popsány, což ji poskytuje přidanou hodnotu z pohledu metodologie. Výsledky této práce poskytují nové poznatky o spojitosti užívaných příjmení a územím se specifickým nářečím. Tím otevírá možnosti pro další výzkum.

Tato práce přináší nový pohled na studium jazykových a geografických aspektů příjmení v České republice a poskytuje kvalitní základ dalšímu výzkumu prostorové distribuce příjmení a nářečních variant. Výsledky a metodika této práce mají potenciál uplatnění v dalších výzkumech v oblastech dialektologie, demografie, historie a jazykového zeměpisu. Demonstruje také užitečnost využití moderních geoinformačních a analytických nástrojů pro zvýšení efektivity výzkumu.

1 CÍLE PRÁCE

Hlavním cílem diplomové práce je zjistit prostorovou podmíněnost mezi vybranými příjmeními a jejich nářečními variantami na území Česka. Na základě získaných dat prostorovými analýzami a mírou výskytu potvrdit či vyvrátit hypotézu, že bydliště osob se specifickými příjmeními korelují s nářečními oblastmi českého jazyka.

Pro dosažení cíle byly stanoveny následující dílčí cíle:

1. S pomocí dialektologů z Ústavu pro jazyk český Akademie věd České republiky vybrat slova, která tvoří základ specifických příjmení.
2. Připravit dvě sady dat:
    a. vrstvy prostorové distribuce osob s vybranými příjmeními z dat Ministerstva vnitra České republiky publikované na webu KdeJsme.cz
    b. digitalizované mapy nářečních variant slov tvořících vybraná příjmení ze zdroje Český jazykový atlas
3. Stanovit vhodné metriky pro míru výskytu jednotlivých příjmeních a pomocí nich určit míru geografické shody mezi příjmením a nářeční variantou.
4. Na základě výsledků prostorových analýz a hodnot nekvantifikované i kvantifikované míry geografické shody potvrdit či vyvrátit hypotézu, že bydliště osob se specifickými příjmeními korelují s nářečními oblastmi českého jazyka.

Práce má díky úzké spolupráci s dialektology z ÚJČ AV interdisciplinární přesah a otevírá tak nové možnosti dalšího a podrobnějšího výzkumu z pohledu dialektologie, demografie, historie nebo jazykové geografie. K tomu mohou přispět nejen výsledky analýz, ale také vytvořené metriky, které jsou v práci využity k interpretaci výsledků prostorových analýz a v neposlední řadě k potvrzení či vyvrácení stanovené hypotézy. Hlavním přínosem této práce v budoucím výzkumu podobného zaměření, je vyvinutá metodika, která má potenciál, být uplatněna v dalších výzkumných projektech, zabývajících se podobnými vztahy mezi jazykem, jmény a prostorem.

2 SOUČASNÝ STAV ŘEŠENÉ PROBLEMATIKY

Znázorňování prostorové distribuce příjmení a jejich nářečních variant představuje interdisciplinární výzvu, která v sobě kombinuje poznatky z oblasti geoinformatiky, dialektologie a kartografie. V této kapitole je provedena rozsáhlá rešerše současné literatury, výzkumů a studií, které se zaměřují na metodologii mapování jazykových fenoménů a jejich praktických aplikací v České republice i ve světě.

Výzkumy zaměřené na analýzu prostorového rozložení jazykových a onomastických prvků odhalují fakt, že geografické a lingvistické hranice často nejsou shodné. To může vést k zajímavým objevům o historických migračních trendech a kulturních výměnách. Touto problematikou se zabývá dílčí jazykovědní disciplína jazykový zeměpis neboli geolingvistika (Kloferová, 2017c), která propojuje jazykovědné znalosti dialektologie s geografickými metodami, a pomocí specifických jazykovězeměpisných metod jsou sledované jazykové jevy zakreslovány na mapu. Ty nám umožňují zobrazit prostorovou distribuci jednotlivých nářečí a jazykových jevů. Ucelený soubor takových map se nazývá jazykový atlas (Kloferová, 2017b).

V České republice představuje klíčový zdroj v tomto ohledu Český jazykový atlas (Balhar et al., 2012), který přináší výsledky rozsáhlého přímého výzkumu českých nářečí a běžné mluvy. Atlas významně obohacuje poznání našeho o prostorovou složku, jelikož obsahuje podrobné informace o geografickém rozložení mluveného jazyka na našem území. Publikací, která se věnuje českému jazyku a jejímu vývoji, je časopis Naše řeč vydávaný Ústavem pro jazyk český Akademie věd České republiky. Poskytuje česky a anglicky psané studie o současné češtině a jejím historickém vývoji.

Podobným studovaným dílem (i když nikoliv tak rozsáhlým), je kniha An Atlas of English Dialects, kterou autoři Clive Upton a J. D. A. Widdowson (2006) ve dvou vydáních publikovali v roce 1996 a 2006. Jedná se o soubor 90 map Velké Británie, které zobrazují jednotlivé rozložení dialektů angličtiny napříč celým souostrovím.

Při zkoumání vztahů mezi dialektem a jeho prostorové distribuci a jmény je důležitým bodem pochopení vzniku a významu jednotlivých jmen a příjmení. K tomu slouží subdisciplína lingvistiky označovaná jako onomastika. Je chápána jako nauka o vznikání, fungování a strukturaci propriální sféry jazyka neboli nauka o vlastních jménech (Pleskalová, 2017b).

Geografickou distribucí jmen a příjmení a jejich vztahem s dialektologickými oblastmi se ve Velké Británii věnoval ve svých studiích Paul Longley, který se v různých kolektivech spoluautorů zaměřoval na využití moderních geoinformatických nástrojů (Van Dijk, Longley, 2020a) a statistických metod (Van Dijk, Lansley, Lan, Longley, 2019) pro mapování jmen a příjmení v různých kulturních kontextech (Cheshire, Longley, 2011a). Tyto studie demonstrují, jak moderní pokročilé technologie umožňují hlubší analýzu a pochopení prostorových vzorců a nabízejí nové možnosti pro interpretaci lingvistických dat.

Geografické rozložení jednotlivých příjmení a jejich intenzitu zkoumá Fiona McElduff (2008). Využívá k tomu data z volebních seznamu z roku 2001 pomocí nichž identifikuje, jak se rozložení příjmení liší mezi různými regiony a jak jsou tato data spojena s historickými migracemi a usazením populace.

Výzkumem s výrazným onomastickým přínosem je studie ze Španělska, ve které autoři Pablo Mateos a Ken Tucker (2008) odhalují nerovnoměrné rozdělení nejpopulárnějších příjmení a dávají ho do souvislosti s historickým kontextem a ve své studii dále přináší metody, které kombinují geografickou a statistickou analýzu jmen a jejich původu. To umožňuje hlubší analýzu a pochopení prostorových vzorců a přináší nové možnosti pro

interpretaci lingvistických dat. Své poznatky a výsledky analýz navíc srovnávají s daty z jiných hispánských zemí a odhalují rozdíly mezi nejpopulárnějšími jmény napříč těmito zeměmi.

2.1 Geografická distribuce nářečí

Rozmístění nářečí v prostoru je jednou z důležitých součástí dialektologického výzkumu. Umožňuje analyzovat a lépe porozumět jazykovým rozdílům a sociokulturním dynamikám v různých regionech. V následující části práce je přiblížena teorie, metodologie a publikované výsledky výzkumu geografické distribuce nářečí u nás i ve světě.

Jazykové atlasy

Kloferová (2017b) píše: *„Jazykový atlas je soubor map zobrazujících jazykové jevy v jejich zeměpisném rozšíření v předem vymezené oblasti. Jazykový atlas je završením studia jazyka zkoumaného za využití jazykovězeměpisných metod.“*. Mapy zpravidla doprovází komentáře, které jsou buď součástí svazku s mapami nebo tvoří samostatný svazek.

Jazykové atlasy mohou být rozděleny dle způsobu zpracování a výkladu a podle povahy mapovaných jevů nebo podle rozsahu zkoumaného území. V kontextu práce je důležitější rozdělení podle zkoumaného rozsahu. Jazykové atlasy dělíme dle Kloferové (2017b) na:

1. nadnárodní – větší území pokrývající více států (např.: *Atlas linguarum Europae*)
2. národní – zobrazuje území jednoho státu (např.: *Český jazykový atlas*)
3. regionální – zobrazuje vybraný region na území jednoho nebo více států (např.: *Atlas językowy Śląska*)

Jazykový atlas může být podkladem pro další výzkumy nejen v oblasti lingvistiky. Mohou také sloužit jako vzdělávací materiál pro lepší pochopení vztahů jazyka a prostoru.

První dílo považované za jazykový atlas vytvořil George Wenker s kolegy v Německu na konci 19. století. Hlavním zdrojem dat byl rozsáhlý poštovní průzkum dialektů napříč regionem. Tyto dotazníky byly zaslány 40 učitelům, kteří byli požádáni o přeložení předpřipravených vět do místního dialektu (Boberg, Nerbonne, Watt, 2018).

Český jazykový atlas

Klíčovým zdrojem pro pochopení geografické distribuce českých nářečí a jedním z důležitých zdrojů celé práce je Český jazykový atlas. Jedná se o rozsáhlé šestidílné dílo, které bylo sestaveno týmem lingvistů a dialektologů z Ústavu pro jazyk český Akademie věd České republiky pod vedením Jana Balhara (1. – 6. díl) a Pavla Jančáka (1. – 3. díl). V tištěné podobě byl postupně vydáván mezi lety 1992 a 2011. Elektronická kompletní verze atlasu je od roku 2012 dostupná na webových stránkách v podobě PDF (Balhar et al., 2012) a od roku 2018 je atlas kompletně dostupný ve formě HTML stránky (Balhar et al., 2018a).

Dle Kloferové (2017b) podává ČJA *„nejzevrubnější jazykovězeměpisnou analýzu českého národního jazyka“* a označuje ho za stěžejní dílo české dialektologie. Jedná se o velice rozsáhlé dílo, které se skládá ze 6 svazků:

1. lexikum spojené s tradičním způsobem života (místní prostředí, domácí prostředí, člověk)
2. lexikum spojené s tradičním způsobem života (zahrada a sad, živočišstvo, les a rostlinstvo, krajina, čas a počasí, vesnice dříve a nyní)
3. lexikum spojené s tradičním způsobem života (polní zemědělské práce, hospodářská usedlost, zemědělské nářadí a nástroje, dobytek, drůbež)
4. morfologie
5. hláskosloví a syntax

6. dodatky

Atlas celkem obsahuje 1558 map, které dokumentují stav běžné české mluvy a nářečí v 60. – 70. letech 20. století, ve kterých probíhal terénní dialektologický výzkum. Ten se uskutečnil pomocí dotazníkového šetření na základě *Dotazníku pro výzkum českých nářečí* a obsahoval celkem 2649 položek. Výzkum proběhl ve 420 venkovských obcích tradičního osídlení a 57 měst na českém jazykovém území. Sesbíraný materiál představuje jazykové jevy ve vývojové, geografické i generační úplnosti v rozmezí posledních sta let. Dokumentován byl i stav českého jazyka ve několika zahraničních lokalitách, kde jsou doloženy enklávy českého jazyka – Polsko, Srbsko, Chorvatsko, Bosna a Hercegovina a Rumunsko (Kloferová, 2017b).

Atlas navazuje na dlouhou tradici české dialektologie, jejíž počátky sahají až do roku 1864 k dílu Aloise Vojtěcha Šembery *Základová dialektologie československá,* jež je považováno za *„zakladatelské dílo české a slovenské dialektologie“* (Balhar et al., 2012). Zásluhou Františka Bartoše pak byla v roce 1886 a 1895 zachycena nářeční situace na Moravě a ve Slezsku v díle *Dialektologie moravská.* Systémový výzkum českých nářečí byl nejaktivněji rozvíjen po druhé světové válce a přinesl významné metodologické inovace a rozšířené výzkumné techniky.

ČJA je klíčovým zdrojem pro studium českých nářečí, poskytuje unikátní pohled na jazykovou variabilitu českého jazyka a je důležitý pro pochopení historických a sociolingvistických procesů v české společnosti. Je navržen tak, aby byl přístupný nejen odborníkům, ale i široké veřejnosti. Tím tak přispívá nejen jako vědecké dílo, ale také jako vzdělávací nástroj sloužící k rozšíření vědomostí o českém jazyce a k širšímu pochopení jazykové rozmanitosti České republiky.

An Atlas of English Dialects

Clive Upton a J. D. A. Widdowson (2006) v tomto atlase přináší rozsáhlý pohled na regionální dialekty angličtiny ve Velké Británii. Dílo je založeno na výzkumu Survey of English Dialects, což je považováno za nejrozsáhlejší záznam o regionálním mluveném jazyce v Anglii. Dílo obsahuje celkem 90 map, které zobrazují geografické rozdělení zkoumaných jazykových jevů.

Sběr dat pro tento atlas zahrnoval dotazníkové šetření, fonetické záznamy a magnetofonové nahrávky. Zaměřen byl na starší, místně narozené osoby s nízkým vzděláním, čímž bylo zajištěno, že jejich mluva nebyla výrazně ovlivněna vnějšími sociálními vlivy nebo moderními technologiemi.

Atlas silně poukazuje na fakt, že britská angličtina je a vždy byla jazykem mnoha rozdílných dialektů. Tyto dialekty se liší nejen slovní zásobou, ale také gramatikou nebo výslovností. V knize je také zmíněn historický kontext a autoři poukazují na to, že se dialekty vyvíjely také v důsledku historických událostí, které na území Velké Británie probíhali. Výraznými událostmi, které silně ovlivnily a do jisté míry formovaly dialekty angličtiny ve Velké Británii jsou například invaze Vikingů a Normanů, které probíhali od 9. století.

V závěru atlas odhaluje, že ačkoliv rozmach moderních technologií v současné společnosti vedou k určité standardizaci jazyka, britská angličtina si stále uchovává svoji výraznou rozmanitost dialektů. Autoři zde zdůrazňují, že právě tato diverzita je klíčovou součástí kulturního dědictví a reflektuje bohatou historii regionálních identit formovaných mimo jiné právě dialekty.

Atlas of English Dialects je dílem, které slouží jako akademický zdroj, ale také jako výzva k dalšímu zkoumání dialektů a možná také trochu výzvou k zachování a uchovávání dialektické rozmanitosti anglického jazyka ve Velké Británii.

2.2 Geografická distribuce jmen a příjmení

Rozšíření jmen a příjmení v různých oblastech reflektují současné dialektologické a onomastické studie a výzkumy jako významný indikátor historických migrací, sociálních změn a kulturních interakcí. Při tvorbě této práce byly prostudovány vhodné materiály, které se zabývají prostorovou koncentrací a distribucí jednotlivých jmen a příjmení.

Geografickou distribuci příjmení ve Španělsku v roce 2004 zkoumali Pablo Mateos a Ken Tucker (2008) ve studii frekvencí křestních jmen a příjmení podle telefonních seznamů. Zatímco rozdělení křestních jmen ve španělské populaci odpovídá vzoru jako v jiných zemích, které Tucker studoval, u příjmení bylo zjištěno jedinečné rozdělení s mnohem vyšší koncentrací populace v nejpopulárnějších příjmeních než v jiných zemích.

Vysvětlení těchto zjištění autoři hledají ve třech kombinovaných procesech – křesťanské osidlování Iberského poloostrova od raného středověku, proces nucené změny jmen způsobené španělskou inkvizicí[1] a také fenoménem vyšší frekvence *inbreedingu příjmení*[2], který označuje sňatek mezi jedinci stejného příjmení.

Analýza skupin jmen klasifikovaných podle jejich jazykového původu naznačuje potenciál použití kvantitativní analýzy geografické distribuce jmen k odhalení historických procesů osídlení a migrace. Studie tak odhaluje původní jazykové regiony ve Španělsku ve středověku a to, jak tyt regiony stále strukturují populaci do dnešní doby. Frekvence výskytu nejpopulárnějších příjmení ve Španělsku autoři srovnávají s frekvencemi s dalšími čtyřmi španělsky mluvících zemích nebo v zemích s početnou španělsky mluvící menšinou (Argentina, Mexiko, Venezuela a USA). Díky tomuto srovnání zjišťují, jak se různé populace usazovaly v čase a jak na rozmanitost příjmení působily nucené pojmenovávací praktiky v bývalých koloniích (Mateos, Tucker, 2008).

Podrobně studováno bylo dílo Fiony McElduff a kolektivu (2008), ve kterém je zkoumána frekvence a geografické rozložení příjmení ve Velké Británii založené na datech z volebního seznamu z roku 2001. Studie se zaměřuje na identifikaci regionálních rozdílů v rozložení příjmení a snaží se propojit tyto rozdíly s historickými a migračními trendy a sociekonomickými charakteristikami jednotlivých regionů.

Výzkum poukazuje na to, jak příjmení odrážejí historické migrace a usazování obyvatelstva ve Velké Británii. Některé regiony Velké Británie, jako je Wales nebo Cornwall, vykazují vysokou míru koncentrace specifických příjmení, což odráží nižší míru migrace a větší izolaci těchto oblastí v historickém kontextu. Naopak v Londýně a dalších větších urbanizovaných oblastech v jihovýchodní Anglii je zaznamenána vysoká diverzita příjmení. To koresponduje s historickými i současnými vlnami imigrace a vysokou úrovní demografické dynamiky.

Ke kvantifikaci diverzity příjmení byl použit index Yuleovo K, který umožňuje srovnání míry unikátnosti příjmení mezi různými regiony a identifikuje oblasti s vysokou mírou endogamie či exogamie.

Závěry výzkumu ukazují, že informace o distribuci příjmení jsou zásadním zdrojem pro analýzu a pochopení demografických a sociálních změn v prostoru. Příjmení odráží nejen historickou migraci a socioekonomické podmínky regionů, ale také mohou sloužit jako indikátory pro studium současných demografických trendů. Studie nabízí nový pohled na využití demografických dat pro hlubší analýzu sociálně-geografických procesů ve Velké Británii (McElduff a kol., 2008).

---

[1] právní instituce španělských panovníků, která udržovala katolickou ortodoxii a bojovat s kacířstvím (ve Španělsku vznikla v roce 1478). (Wikipedia, 2024)

[2] z anglického *surname inbreeding* (Mateos, Tucker, 2008)

Zajímavý pohled na využití dat o geografické distribuci příjmení nabízí výzkum autorů J. Cheshireho a P. Longleyho (2011b) publikovaný v akademickém žurnálu *Procedia – Social and Behavioral Sciences*. Tato studie se zaměřuje na identifikaci geografických koncentrací příjmení s cílem poskytnout lepší porozumění jejich historickému původu, současnému rozšíření a vztahům s dalšími příjmeními a místními názvy.

Hlavním cílem studie je demonstrovat, jak mohou být příjmení využita jako efektivní nástroj pro studium demografických změn. Příjmení poskytují jedinečný pohled na historické osídlení a migrace a mohou pomoci identifikovat jak dlouhodobé, tak nedávné populační pohyby.

Využitím metodiky založené na odhadech hustoty jádra – Kernel Density Estimation (KDE) je výzkum schopen identifikovat oblasti s nejvyšší koncentrací jednotlivých příjmení a analyzovat jejich geografickou distribuci v národním, regionálním a lokálním měřítku. Tato metoda umožňuje autorům generovat spojité povrchy hustoty z diskrétních bodových dat. KDE je upravena tak, aby reflektovala různorodé geografické charakteristiky příjemní od velmi lokalizovaných až po široce rozptýlené.

Studie dospěla k závěru, že příjmení ve Velké Británii jsou výrazně nenáhodně rozložena a tendují ke shlukování kolem svých historických oblastí původu. Tyto výsledky potvrzují, že příjmení mohou sloužit jako kulturní identifikátory, které poskytují bohaté kulturní informace. Studie také odhaluje, že moderní distribuce příjmení může být ovlivněna nedávnými migračními událostmi, což má dopady na národní, regionální a lokální populační struktury (Cheshire, Longley, 2011b).

S průlomovým přístupem k analýze demografických změn a migračních vzorců ve Velké Británii na základě geografické distribuce příjmení, přišli Justin Van Dijk a Paul Longley (2020b), kteří se ve svém výzkumu zaměřili na kombinaci historických a současných dat ke sledování změn v populaci na základě příjmení od roku 1881 do současnosti (srovnání je zobrazeno pro roky 1998 a 2016). Výzkum sleduje geografický původ a distribuci celkem 59 218 příjmení.

Cílem výzkumu je vytvoření placiálních[3] profilů příjmení, které ukazují, jak se specifická příjmení geograficky a časově rozložila ve Velké Británii. Studie se snaží mapovat, jak se rodinná jména přenášejí mezi generacemi a jak se mění jejich geografická distribuce. To poskytuje důležité poznatky o historických migračních trendech a demografické stálosti.

Zdrojem dat byla historická data sčítání a současné populační registry. Pro jejich analýzu byla využita metoda KDE k odhadu hustoty rozložení příjmení. To umožnilo autorům identifikovat a vizualizovat geografické vzorce spojené s příjmeními bez omezení způsobených změnami administrativních hranic.

Studie dokazuje, že analýza příjmení může efektivně odhalovat migrační a demografické trendy i na národní úrovni. Výsledky ukazují, že velké městské aglomerace jako Londýn a Birmingham mají vyšší diverzitu v původu příjmení, což odráží vyšší míru migrace a sociální mobility. Na druhé straně menší a stabilnější oblasti vykazují větší koncentraci místně specifických příjmení (Van Dijk, Longley 2020b).

[3] z anglického *platial geo-temporal demographic (*tj. demografická analýza založena na specifických místech a časových intervalech) (Van Dijk, Longley, 2011b)

2.3 Metodika

Výzkum geografické distribuce nářečí, jmen a příjmení vyžaduje použití vhodných statistických metod pro analýzu těchto jevů. Záleží také na volbě vhodných a co nejpřesnějších zdrojů dat. Správný výběr metod a zdrojových dat je klíčový pro interpretaci konečných výsledků zkoumání. Vhodný výběr umožňuje nejenom identifikaci prostorových vzorců, ale i hlubší pochopení sociolingvistických a demografických procesů, které často formují celé regiony a jejich identitu. V této části práce přibližuje některé použité metody a zdroje dat v současných vědeckých výzkumech v oboru demografie, dialektologie a geolingvistiky.

2.4 Zdroje dat

Významnou proměnnou, která ovlivňuje závěry každého výzkumu jsou data, která do analýz vstupují. Zdroje dat pro analýzy geografické distribuce nářečních jevů a onomických objektů[4] jsou různorodé.

Tradičně se dialektologický výzkum opíral o terénní průzkumy a sběr dat prostřednictvím dotazníků a rozhovorů což ilustruje například *Český jazykový atlas*, ve kterém byl zdrojem dat terénní výzkum, což umožnilo zachycení autentických jazykových projevů v různých oblastech (Balhar et al., 2012). Terénní výzkum a sběr dat prostřednictvím dotazníků a rozhovorů byl klíčový při tvorbě jazykových atlasů v 19. a 20. století (Boberg, Nerbonne, Wat, 2018).

V posledních dekádách se ovšem jazykový výzkum do značné míry transformoval díky přístupu ke komplexním elektronickým databázím. Trend využívání dostupných komerčních a administrativních záznamů demonstrují Mateos a Tucker (2008), kteří pro svou studii využili data ze španělského telefonního seznamu, což jim umožnilo analyzovat rozložení příjmení. Jako zdroj dat o geografické distribuci jmen a příjmení využila ve své studii Fiona McElduff s kolektivem (2008) volební seznamy. Jednalo se o volební seznamy z roku 2001, které obsahovaly data o jménech, příjmeních a adresách všech osob starších 16 let, kteří byli oprávněni hlasovat ve volbách ve Velké Británii. Spolehlivým zdrojem dat jsou také národní sčítací data (historická i současná), která vznikají při sčítaní lidu. Ta používají ve své studii Van Dijk a Lansley (2019). Podobným příkladem je použití kombinace dat z britského sčítání lidu a volebních seznamů ve studii Jamese Cheshira a Paula Longleyho (2011a). To jim poskytlo detailní pohled na geografické vzorce v distribuci příjmení ve Velké Británii.

Zdroje dat pro výzkum jazykové variability a onomastiky se v průběhu let a vyvíjeli od manuálního sběru informací až po sofistikované analýzy založené na rozsáhlých elektronických databázích. Tento vývoj vědcům umožňuje provádět složité analýzy a interpretace, které byly dříve omezeny, kvůli nedostatku široce a jednoduše dostupných dat.

2.5 Analytické metody

Ve studovaných materiálech došlo k využití široké škály statistických a analytických metod zvolených pro studium geografické distribuce příjmení a nářečí. Tyto metody poskytují této práci náhled do různorodých přístupů k výzkumu a ukazují, jak lze efektivně interpretovat

[4] dle Pleskalové (2017) jde o *„takové objekty, které je třeba z komunikačních důvodů pojmenovat vlastním jménem“*

a vizualizovat rozsáhlé datasety pro odhalení vzorů v jazykové a geografické distribuci zkoumaných jevů.

Nejčastěji se ve studovaných materiálech vyskytuje použití statistické metody Kernel Density Estimation (KDE). Jedná se o odhad hustoty jádrovým vyhlazováním. Metoda používá jádro (váhová funkce), které přiřazuje váhy blízkým datovým bodům. Metodu aplikují ve své práci Van Dijk a Lansley (2019), kteří ji využívají k odhadu hustoty bodových dat získaných z historických sčítání lidu v kombinaci se současnými registry. Tato metoda umožňuje identifikaci a vizualizaci geografických vzorců bez omezení způsobené změnami administrativních hranic. Z tohoto důvodu ji využívají také Van Dijk a Longley (2020b), kteří tuto metodu aplikují na 1,2 milionu příjmení, což jim umožňuje sledovat migraci a demografické změny ve Velké Británii od historie až po současnost i přes změny administrativních hranic. Metoda umožňuje generovat z diskrétních bodových dat spojité povrchy hustoty, což poskytuje detailnější pohled na prostorové vzorce příjmení. Upravenou metodu KDE používají Cheshire a Longley (2011). Ti si parametry metody upravili tak, aby reflektovala různorodé geografické charakteristiky příjmení od velmi lokalizovaných až po široce rozptýlené.

Fiona McElduff a její kolektiv (2008) využívají k analýze dat metodu Yuleova indexu K pro hodnocení diverzity příjmení. Tento statistický nástroj umožňuje srovnání míry unikátnosti příjmení mezi různými regiony a identifikuje tak regiony s vysokou mírou převážně přistěhovalecké endogamie[5] či exogamie[6]. Tato metoda nám pomáhá, podobně jako KDE, identifikovat nejen základní distribuci, ale i sofistikovanější vzorce spojené s demografickými a historickými změnami v prostoru.

Kombinaci několika analytických metod k prozkoumání distribuce jmen a příjmení používají ve své studii Mateos a Tucker (2008). Ti nejdříve analyzují frekvenci jmen pomocí explorativní geografické analýzy, pomocí které odhalují geografické vzorce v distribuci příjmení. Dále analyzují frekvenci jmen a příjmení z dat telefonních seznamů. Pro analýzu rozložení jmen podle jejich frekvence autoři použili Zipfovo rozdělení. Jedná se o statistické rozdělení, jehož základní princip spočívá v tom, že frekvence výskytu slova ve velkém textu je nepřímo úměrná jeho pořadí dle četnosti – druhé nejčastější slovo se vyskytuje s přibližně poloviční frekvencí slova nejčastějšího, třetí nejčastější s jednou třetinou frekvence slova nejčastějšího atd (Příručka ČNK, 2013).

Výzkum geografické distribuce dialektu a příjmení nabízí využití širokého spektra analytických a statistických metod a postupů. Pro každé studované téma a území je potřeba vybrat ideální metodu a správně ji aplikovat. Metoda KDE, která je ve studované literatuře nejfrekventovaněji využívána, je výborná metoda pro analýzu diskrétních bodových dat, které nejsou vázány stálými administrativními hranicemi a výzkum potřebuje data od těchto hranic „odvázat“. Použití Yuleova indexu je zase vhodné pro studium diverzity nejen jazykových jevů v určitém území. Je možné pomocí něj porovnávat různé regiony a oblasti dle jejich míry diverzity a odhalovat tak zajímavé poznatky v oblasti demografie. Využití statistických zákonů a pravidel rozdělení je zase vhodné pro identifikaci nejfrekventovaněji se vyskytujících jazykových jevů (příjmení) v určitém stanoveném prostoru.

[5] uzavírání sňatků mezi členy téže skupiny (sociální, etnické, územní atd.), (Justoň, 2017b)

[6] uzavírání sňatků mimo vlastní skupinu (Justoň, 2017a)

2.6 Technologické přístupy

V dnešní době moderních technologií se i dialektologické a geolingvistické výzkumy spoléhají na pokročilé technologie a sofistikované analytické metody, které nám využití technologií poskytuje. Technologie tak poskytují nejen přesnější nástroje na analýzu, čímž přispívají k lepším a detailnějším výsledkům studií, ale také zpřístupňují výsledky široké veřejnosti prostřednictvím interaktivních a vizuálně atraktivních prezentací.

Detailní vizualizace jazykových procesů v průběhu času umožňuje již dříve zmíněná statistická metoda KDE, díky které lze za pomocí geoinformačních systémů přehledně vizualizovat, jak se příjmení v průběhu času rozšiřují a koncentrují do určitých oblastí. Sílu tohoto přístupu v kombinaci s atraktivní vizualizací demonstrují Van Dijk a Longley (2020b), kteří k prezentaci výsledků studie využívají interaktivních webových map, které ukazují, jak se geografické rozložení příjmení mění v čase.

Stejní autoři představují metodologii pro kalkulaci a vizualizaci distribuci příjmení v historickém i současném kontextu ve Velké Británii. Klíčovým přínosem této studie je vývoj interaktivní platformy, která veřejnosti umožňuje přístup k detailním statistikám jednotlivých příjmení. Tato platforma poskytuje vizuální reprezentaci geografických a časových trendů, ale také navíc umožňuje uživateli provádět vlastní analýzy (Van Dijk, Longley, 2020a).

Tyto přístupy demonstrují sílu moderních technologií ve výzkumu a zpřístupňují často složité odborné závěry studií široké veřejnosti. Použití interaktivních a vizuálně poutavých metod zároveň usnadňuje šíření poznatků a podporuje další výzkum. Umožňují nám lépe pochopit, jak studium příjmení a nářečí obyvatel může přinést lepší porozumění historickým událostem a tomu, jak se tyto události podepsali na dnešní sociokulturní dynamice našich společností.

3 METODY A POSTUP ZPRACOVÁNÍ

Tato kapitola je zaměřena na popis metodologického rámce a použitých technologií, které byly klíčové pro dosažení cílů této diplomové práce. Práce se věnuje analýze míry geografické shody mezi příjmeními a dialekty v České republice, což vyžadovalo komplexní přístup ke zpracování a analýze rozsáhlých datových sad. Důležitou součástí tohoto procesu bylo využití pokročilých nástrojů GIS, digitalizačních technik a statistických metod, které umožnily efektivní manipulaci s daty a jejich následnou vizualizaci.

Kapitola poskytuje detailní přehled o všech technologiích a postupech, které byly použity k dosažení výzkumných cílů. Úvodní část se věnuje popisu zdrojů, ze kterých byly připraveny podkladová data pro analýzy. Dále se věnuje metodám práce s daty z těchto zdrojů, včetně Českého jazykového atlasu a webové aplikace KdeJsme.cz. Popsány jsou přístupy k digitalizaci a vektorizaci rastrových dat, identifikaci oblastí nevhodných pro analýzu a studium literatury.

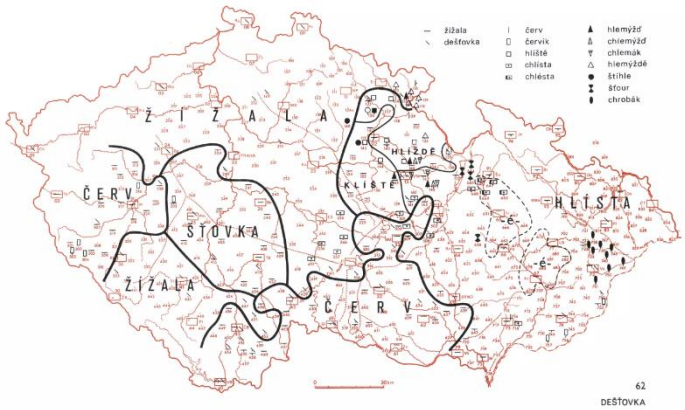
V dalších částech se kapitola věnuje popisu statistický nástrojů a analytických technik, které byly aplikovány pro kvantifikaci a interpretaci výsledků. Kapitola detailně popisuje metodiku, která byla vyvinuta speciálně pro potřeby této práce, ve spolupráci s odborníky z Českého jazykového ústavu Akademie věd ČR. Celkově je tato kapitola určena k poskytnutí jasného a strukturovaného základu pro pochopení metodického přístupu, jehož výsledky jsou prezentovány v následujících částech práce.

3.1 Zdroje dat

Datové podklady pro tuto práci byly převzaty ze dvou hlavních zdrojů dat – Českého jazykového atlasu a webové aplikace KdeJsme.cz, která přebírá data od Ministerstva vnitra ČR. Pro statistické analýzy byla využita data Českého statistického úřadu, která jsou součástí vektorové databáze ArcČR®500 od společnosti ARCDATA PRAHA (2022).

3.1.1 Český jazykový atlas

Je zásadním zdrojem celé práce a poskytuje data o dialektických oblastech a jevech na území České republiky. Data, která jsou v atlase zobrazena v mapách vznikla provedením dialektologického výzkumu na našem území. Použita byla digitální verze ČJA (Balhar et al., 2018a) což usnadnilo celý proces digitalizace a omezilo ho pouze na vektorizaci těchto dat. V případě použití tištěné verze by bylo nutné data digitalizovat do digitálního formátu.



Obr. 1 - ukázka mapy z Českého jazykové atlasu (zdroj: Český jazykový atlas)

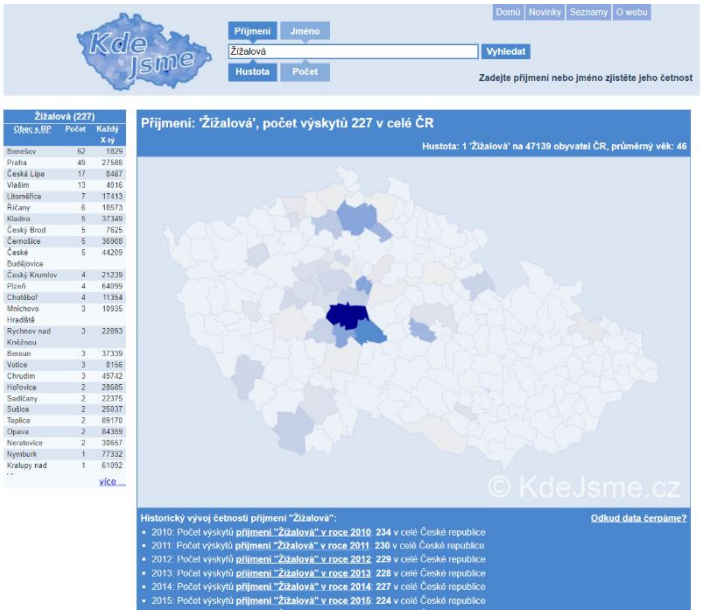
3.1.2 KdeJsme.cz

Jedná se o jednoduchou webovou aplikaci jejímž autorem je Ondřej Malačka (2011). Aplikace umožňuje v jednoduchém rozhraní vyhledávat data o četnosti vybraného příjmení a zobrazuje data v podobě mapy a tabulky. Aplikace zobrazuje data pro mužské a ženské varianty zvlášť, je proto nutné vyhledávat přesné znění příjmení i v jeho přechýlené variantě.

V mapě jsou data zobrazena jako hustota počtu nositelů hledaného příjmení v jednotlivých ORP na celkový počet obyvatel v ORP. Hustotu v mapě reprezentují stupně barvy od nejtmavší po světlejší. Vedle map je umístěna tabulka, která poskytuje informace o názvu jednotlivých ORP a počtu nositelů příjmení v nich. Stejně jako v mapě i zde je uveden údaj o hustotě výskytu příjmení. Stejná data lze zobrazit pro každé ORP jednoduchým kliknutím na vybrané ORP. Aplikace dále poskytuje informaci o celkovém počtu nositelů příjmení v ČR, hustotu na celkový počet obyvatel ČR a průměrný věk nositelů. K údajům o celkovém počtu příjmení a průměrném věku jeho nositelů poskytuje web také data historického vývoje těchto ukazatelů, a to od roku 2010.

Data o jménech a příjmeních zobrazené v aplikaci jsou interpretací volně dostupných dat poskytovaných Ministerstvem vnitra České republiky (dále MVČR). Četnost jmen a příjmení generovalo MVČR z agendového informačního systému evidence obyvatel. Pro tuto činnost již ovšem dále nedisponuje zákonnou oporou. Od roku 2018 totiž platí nařízení Evropského parlament a Rady EU 2016/679 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů (MVČR, 2024). Toto nařízení znemožňuje MVČR tyto data dále uchovávat a zveřejňovat. Institucí, která je oprávněna tato data sbírat a poskytovat z nich statistické údaje je dle § 18 odst. 1 písm. b) a c) zákona č. 89/1995 Sb., o státní statistické službě pouze Český statistický úřad (ČSÚ, 2024). Z tohoto důvodu byla také k 1. 1. 2020 ukončena aktualizace webové aplikace.

Aplikace tak nabízí nejaktuálnější veřejně dostupná data, která odpovídají stavu v roce 2016. Tento fakt ovšem nijak nesnižuje kvalitu výsledků této práce, jelikož dynamika změn příjmení napříč celou společností probíhá velice pomalu a velké změny, které by výrazně ovlivnily výsledky výzkumu, se neprojevují v rámci několika jednotek let.



Obr. 2 - webová aplikace KdeJsme.cz (screenshot)

3.1.3 ArcČR®500

Hlavním zdrojem vektorových geografických dat je digitální vektorová geodatabáze ArcČR®500 ve verzi 4.1. Tu poskytuje jako volně dostupnou společnost ARCDATA PRAHA. Verze 4.1 byla vydána v červnu 2022 a obsahuje databázi Administrativního členění České republiky odvozenou od z databáze RÚIAN, která je obohacena o vybrané statistické údaje z ČSÚ. Ty zahrnují také data z výsledků SLDB 2021 za všechny úrovně administrativního členění (ARCDATA PRAHA, 2022).

Pro potřeby práce jsou z této databáze použity vrstvy administrativních hranic ORP a státní hranice České republiky. Pro analýzy jsou použity data počtu obyvatel ze SLDB 2021 na úrovni ORP.

3.2 Metody práce

Data z hlavních zdrojů byla převedena do prostředí ArcGIS Pro, kde byla následně zpracována. Z těchto dat byly vytvořeny dvě datové sady – ***CJA*** a ***KDE***. Tyto datasety jsou uloženy v geodatabázi práce. Data z těchto datových sad byly následně analyzovány. Pro interpretaci výsledků byly stanoveny nové metriky, které kombinují kvalitativní a kvantitativní statistické metody. Na základě těchto metrik bylo možné vyhodnotit a interpretovat výsledky provedených analýz a potvrdit či vyvrátit hypotézu stanovenou při zadávání práce.

3.2.1 Výběr příjmení

Příjmení byla vybrána na základě výběru jednotlivých slov, které tvoří základy každého z vybraných příjmení. Tento výběr byl proveden ve spolupráci s dialektoložkou PhDr. Martinou Ireinovou, Ph.D. z Dialektologického oddělení Ústavu pro jazyk český Akademie věd České republiky se sídlem v Brně. Celkem bylo vybráno 15 slov, které tvoří základ pro celkem 35 zkoumaných příjmení. Samotná slova byla doplněna o 3 oblasti vymezené izoglosami[7], které charakterizuje specifický nářeční jev.

3.2.2 Příprava dat

První sada dat (***CJA***) pro vybraná slova a příjmení byla vytvořena z digitální verze ČJA převedením do prostředí ArcGIS Pro. K tomu byl využit proces vektorizace. Prvním krokem procesu bylo vytvoření snímků z digitálního ČJA pro mapu ke každému vybranému jevu. Snímky byly pořízeny nástrojem *Výstřižky*, který je integrovaný v operačním systému Windows. Pořízené snímky byly uloženy ve formátu PNG, který poskytuje vysokou kvalitu snímku bez ztráty a komprese dat. Také poskytuje vysokou kvalitu dat při zvětšení či zmenšení snímku z originální velikosti obrázku (ADOBE, 2024). Z těchto důvodů byl tento formát zvolen jako ideální pro potřeby tohoto procesu.

Po převedení map do vhodného formátu bylo dalším krokem vektorizačního procesu georeferencování pořízených snímků map k vektorovým datům obsahující geografické souřadnice. Georeferencování je proces přiřazení skutečných geografických souřadnic k datovým bodům (nejčastěji obrázkům či digitalizovaným mapovým listům) (ESRI, 2024a). Vzhledem k rozdílnému měřítku nebo zobrazované oblasti u některých map, bylo georeferencování provedeno ručně. Data tím nijak zásadně neutrpěla na kvalitě, jelikož šlo pouze o vektorizaci již vizualizovaných dat z Českého jazykového atlasu. Kvalita těchto dat

[7] *„v dialektologii linie na mapě ohraničující území, ve kterém se vyskytuje určitý jazykový jev“* (Kloferová, 2017e)

závisela na kvalitě dialektologického výzkumu a následném kartografickém vyjádření těchto dat, které bylo aplikováno v souladu s odborníky na dialektologickým výzkum.

Vybrané mapy byly georeferencovány ke státním hranicím České republiky z databáze ArcČR®500 ve verzi 4.1 v měřítku 1: 500 000 (ARCDATA PRAHA, 2022). Jako metoda georeferencování byla zvolena metoda *Spline*, která vyžaduje vytvoření minimálně 10 vlícovacích kontrolních bodů a poskytuje jejich vysokou lokální přesnost (ESRI, 2024b). Po vhodném výběru a vložení kontrolních bodů byly data z georeferencovaných map vektorizovány nástrojem *Create Feature*. Pro každou nářeční oblast byla vytvořena vlastní vrstva geometrického typu *polygon*, která byla uložena do předem vytvořeného datasetu v geodatabázi projektu. Ke každé vrstvě byl vytvořen doplňkový informační atribut, který nese informaci o variantě daného jevu. To zjednodušuje vyhledávání a třídění dat v dalších etapách práce. Data byla ukládána do datasetu *CJA* dle systému pojmenování JM*XYz*_CJA_*PRIJMENI* (např.: *JM01a_CJA_ŽÍŽALA*).

Data byla za účelem vyšší výpovědní hodnoty práce zjednodušena s přihlédnutím k celonárodnímu měřítku výzkumu a velikosti zkoumaných administrativních jednotek, kterou stanovila podrobnost zdrojových dat. Z toho důvodu byla vektorizována data pouze pro plošné jevy. Bodové lokální jevy (dublety apod.) nebyly do analýzy zahrnuty, jelikož by při interpretaci nebylo možné určit, zda lokální jev opravdu ovlivňuje nářečí v celém ORP. Důvodem tohoto rozhodnutí je nesourodost velikosti srovnávaných jednotek. Venkovské a městské lokality, které jsou zobrazeny v ČJA bodovým znakem, jsou výsledkem syntézy několika obcí do jedné *výzkumné lokality* (Balhar et al., 2018b). Z toho důvodu není možné zaručit, zda taková okrajová výzkumná lokalita nezahrnuje také obce, které spadají administrativně pod jiné ORP. Nelze také jednoznačně posoudit vliv dané lokality na celkový nářeční trend v celém ORP. Důležitým faktorem tohoto rozhodnutí je také fakt, že v době probíhajícího dialektologického výzkumu se v ČR jako samosprávní jednotky užívaly okresy (MVČR, 2024).

Druhou sadou dat, která byla pro potřeby práce připravena, byla data o geografické distribuci jednotlivých příjmení ***(KDE)***. K tomu byla využita webová aplikace KdeJsme.cz (MALAČKA, 2011), která zobrazuje data o rozmístění příjmení v jednotlivých ORP v České republice.

Prvním krokem bylo získaní dat pro výskyt vybraných příjmení napříč ČR. Jelikož aplikace umožňuje zobrazení dat pouze pro jedno unikátní příjmení, muselo být pro každé příjmení hledání opakováno pro mužskou i ženskou variantu příjmení. V práci se dalé považuje mužská i ženská varianta příjmení za jedno příjmení, jelikož jde pouze o přechýlenou variantu stejného příjmení (např.: Žížala + Žížalová = příjmení „ŽÍŽALA“).

Následně byla vybraná data z této aplikace ručně převedena do programu MS Excel. V něm byla data očištěna o nepotřebné údaje hustoty. Poté byl ke každému názvu ORP přiřazen kód, který odpovídá číselníku RÚIAN (ARCDATA PRAHA, 2022). Přiřazení kódu ke každému záznamu je jedním z klíčových procesů převodu dat do prostředí ArcGIS Pro, jelikož na základě těchto kódů budou data přiřazena k odpovídajícím vrstvám ORP z geografické databáze ArcČR®500 (viz 3.1.3).

Bylo testováno několik pokusů o automatizaci celého procesu přiřazování kódů ORP k odpovídajícím názvům v pořízených datech, což by výrazně zjednodušilo a urychlilo celý proces. Na základě nepřesných a nespolehlivých výsledků těchto testován, byly kódy přiřazeny ručně pro každé příjmení. Takto připravená data pro každé ze studovaných příjmení byla následně uložena ve formátu *.xlsx*, ve kterém je možné data převádět do prostředí ArcGIS Pro.

Samotný převod proběhl pomocí funkce *Add Join,* která přiřadila informaci o četnosti příjmení k jednotlivým ORP, kde byl výskyt příjmení zaznamenán. Toto přiřazení proběhlo na základě jednoznačného identifikátoru – kódu ORP. Data byla přiřazena k vrstvě z geografické databáze ArcČR®500 nesoucí geografické souřadnice jednotlivých ORP a atributů se statickými informacemi (viz 3.1.3.). Pro kontrolu správnosti propojení dat byla data po nahrání porovnána s daty webové aplikace KdeJsme.cz, aby byla ověřena správnost tohoto postupu.

Datová sada *KDE* tak obsahuje data o geografické distribuci všech vybraných příjmení (viz 3.2.1). Každá vrstva nese informace o počtu obyvatel s mužskou i přechýlenou ženskou variantou v jednotlivých ORP, název a kód ORP a v neposlední řadě údaj o počtu obyvatel na úrovni ORP ze SLDB 2021 (viz 3.1.3. Vrstvy byly uloženy do datasetu nástrojem *Feature Class to Geodatabase.* Pro ukládání byl použit systém pojmenování JM*XYz*_KDE_*PŘÍJMENÍ* (např.: JM01a_KDE_ŽÍŽALA).

Při přípravě dat byly vytvořeny dvě datové sady, které byly pojmenovány CJA a KDE. Datová sada CJA ukládá data z celkem 17 digitalizovaných a vektorizovaných map z ČJA, které zobrazují vybrané dialektologické jevy a jejich prostorové rozložení v ČR. Tyto jevy tvoří základ vybraných studovaných příjmení. V datové sadě KDE jsou uložena data o výskytu 35 příjmení v ORP v ČR. Každá vrstva z obou datových sad byla vytvořena na datech z databáze ArcČR®500, kterou poskytuje firma ARCDATA PRAHA a nesou v sobě kromě geografických souřadnic také statistické informace z ČSÚ (ARCDATA PRAHA, 2022).

3.2.3 Zpracování dat

Po procesu přípravy dat bylo nezbytné data dále zpracovat pro nadcházející analýzy. Byly připraveny dvě datové sady, jedna obsahující informace o dialektických jevech a druhá o geografické distribuci příjmení v ORP v České republice.

Zatímco datová sada KDE byla v tuto chvíli připravena k použití, u datové sady KDE byly vyžadovány dodatečné úpravy. V průběhu integrace dat do systému se objevily nečíselné <Null> hodnoty, což komplikovalo zpracování dat.

Pro částečnou automatizaci procesu opravy výše zmíněného problému byl použit skript v programovacím jazyce Python, který nahradil <Null> hodnoty číselnou hodnotou 0 (ESRI, 2023). Skript importuje modul *ArcPy* a nahrazuje hodnoty pomocí funkce *UpdateCursor.* Pro spuštění skriptu byl vytvořen Notebook s názvem „*replace_NULL*“, ze kterého byl následně skript spuštěn.

```
import arcpy
path = r'cesta k vrstvě'
fieldObs = arcpy.ListFields (path)
fieldNames = []
for field in fieldObs:
    fieldNames.append (field.name)
del fieldObs
fieldCount = len (fieldNames)
with arcpy.da.UpdateCursor (path, fieldNames) as curU:
    for row in curU:
        rowU = row
        for field in range (fieldCount):
            if rowU [field] == None:
                rowU [field] = 0
        curU.updateRow (rowU)
del curU
```

Obr. 3 - ukázka kódu pro nahrazení <Null> hodnot v ArcGIS Pro (screenshot)

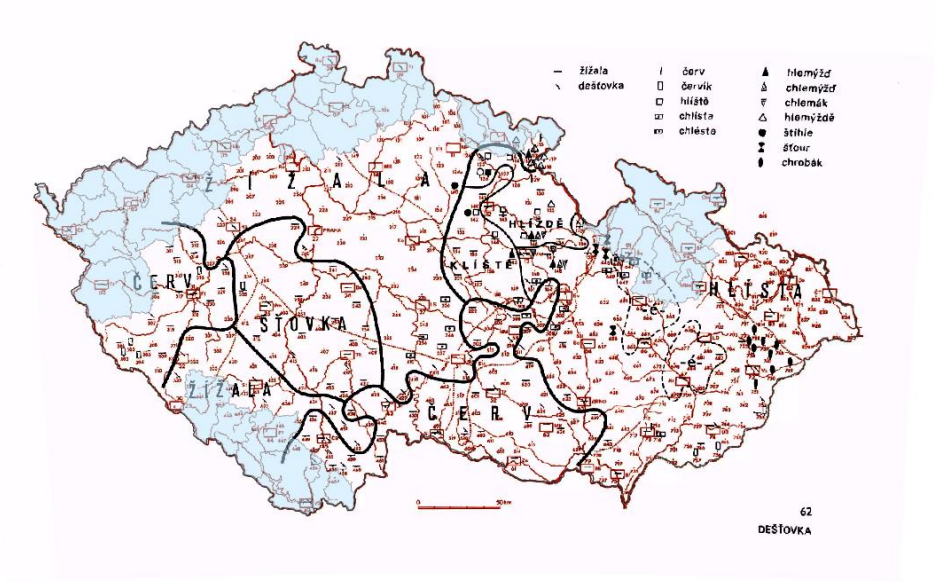
Jelikož práce považuje mužskou a ženskou variantu příjmení za identické příjmení, bylo potřeba sečíst hodnoty pro obě varianty v jednotlivých ORP. V atributové tabulce byl pro tuto potřebu vytvořen nový sloupec „PŘÍJMENÍ_komplet“, ve kterém byl pomocí nástroje *Calculate* Field vypočítán součet četnosti obou variant příjmení v daném ORP.

V posledním kroku byla upravena atributová tabulka tak, aby zobrazovala pouze informace nezbytné pro další analýzu, včetně kódu a názvu ORP, počtu obyvatel v ORP a kompletní četnosti příjmení.

3.2.4 Vymezení pohraničních oblastí

V rámci rozsáhlého dialektologického výzkumu, který probíhal od 60. let 20. století, došlo k pečlivému výběru zkoumaných lokalit a respondentů, čímž byla zajištěna maximální přesnost shromažďovaných dat. Tento proces zahrnoval také vyřazení některých regionů, které nebyly považovány za vhodné pro výzkumné účely. Takovými regiony jsou pohraniční oblasti, kde po druhé světové válce došlo k významným demografickým změnám. Tyto změny zahrnovaly příliv nových osadníků z vnitrozemí, kteří do těchto regionů přinesli rozdílná nářečí a jazykové zvyklosti, což dalo vzniknout novým jazykovým strukturám, odlišným od tradičních dialektů charakteristických pro daný region. Takové oblasti tak nebyly shledány jako vhodné a reprezentativní pro výzkum a byly proto z výzkumu vyřazeny (Balhar et al., 2018b).

Důsledkem vyřazení těchto oblastí je, že pro tyto regiony neexistují podrobná souvislá dialektologická data, což znamená, že neposkytují spolehlivý podklad pro analýzy provedené v této práci. Z toho důvodu bylo nezbytné identifikovat tyto regiony a následně je z dalšího výzkumu vyloučit. K identifikaci byly využity georeferencované mapy z ČJA (viz 3.2.2), které na mapě pomocí bodových znaků zobrazují městské a venkovské výzkumné lokality. Následně byla provedena identifikace ORP, které na svém území neobsahují žádnou výzkumnou lokalitu. Tyto ORP byly označeny jako pohraniční a vyřazeny z konečné analýzy. Spolu s nimi byly vyřazeny i ORP, ve kterých se výzkumné lokality vyskytovali pouze okrajově a nebylo možné s jistotou určit, zda lokality odrážejí nářeční charakter na celém území ORP.



Obr. 4 - identifikované ORP v pohraničních oblastech bez dialektologického výzkumu (zdroj: Český jazykový atlas, ArcČR500)

Vyřazení oblastí s nízkým nebo okrajovým výskytem zkoumaných lokalit bylo nutné pro určení výzkumných jednotek tak, aby bylo možné získat co nejpřesněji interpretovatelná data. Z celkového počtu 206 ORP (včetně Prahy), které se na území ČR nacházejí, bylo 44 z nich identifikováno jako pohraničí bez doloženého dialektologického výzkumu. Do analýzy tedy vstupuje pouze **162 ORP**.

Ve finální vizualizaci analýzy budou zobrazena i pohraniční ORP, přestože byla z některých segmentů výzkumu vyřazena. Důvodem pro jejich zařazení do vizualizace je doložený výskyt některých příjmení v těchto oblastech. Identifikovaná pohraniční ORP byly vyřazeny pouze z analýz, které vyhodnocují míru geografické shody příjmení a nářečí. Toto rozhodnutí bylo přijato za účelem zajištění nejvyšší možné přesnosti a reprezentativnosti výsledků v oblastech, kde je možné na základě dostupných dat určit vztah mezi příjmením a nářečím.

3.2.5 Studium literatury

Ve fázi rešerše byla postupně studována literatura, která se svým odborným zaměřením specializuje na dialektologii, onomastiku a jejich následné propojení s geografií. Studována byla historie dialektologického výzkumu u nás i ve světě, který přinesl mnohé užitečné poznatky o sběru dialektologických dat a jejich zdrojích. Již tyto výzkumy propojovaly sesbíraná data s těmi geografickými, což dalo vzniknout prvním nářečním atlasům.

Studium literatury dále zahrnovalo výzkumy a studie v oblasti jazykové geografie a propojení dialektologie s pokročilými informačními a geoinformačními systémy. To přináší mnohé benefity a zdokonaluje výzkum v oblasti geolingvistiky. Jedná se převážně o studie, které vznikly mimo území České republiky ve Velké Británii a Španělsku. Došlo k seznámení s výsledky těchto studií a důkladnému prozkoumání metod, které byly ve výzkumu použity.

K rešerši studované problematiky byly využity tištěné i elektronické zdroje. K vyhledávání a přístupu k elektronickým zdrojům byly využity platformy jako Google Scholar, Scopus atd.

3.2.6 Vektorizace

V kontextu GIS je jako vektorizace označován proces převodu rastrových dat na data vektorová (ESRI, 2024d). Pro proces vektorizace se využívají tři základní metody (Břehovský, Jedlička, 2024):

- ruční vektorizace
- poloautomatická vektorizace
- automatická vektorizace

Pro vektorizaci dat z ČJA byla zvolena ruční metoda. Při použití této metody je veškerá práce a zodpovědnost při přichytávání jednotlivých vlícovacích bodů na uživateli. U této metody je možné využít asistence počítače pro automatické přichytávání jednotlivých vektorových prvků. V ArcGIS Pro toto umožňuje funkce *snapping*. Z ruční metody se pak stává tzv. čtvrt automatická metoda. Výhodou ruční nebo čtvrtautomatické metody vektorizace je odstranění chyby vytvořené automatizovaným způsobem a nižší nároky na použitý hardware. Naopak nevýhodou může být vyšší časová náročnost celého procesu (Břehovský, Jedlička, 2024).

V práci byla využita čtvrtautomatická metoda vektorizace, jelikož bylo využito možnosti přichytávání bodů pomocí funkce *snapping*.

3.2.7 Odborná konzultace

Jelikož se práce věnuje problematice, která má mezioborový přesah a je založena na dialektologickém výzkumu, je žádoucí, aby byly tato témata konzultována s odborníky v oboru. Veškerá témata spojena s dialektologickým výzkumem byla konzultována s **PhDr. Martinou Ireinovou, Ph.D.** z Dialektologického oddělení ÚJČ AV ČR sídlem v Brně, která je také odbornou konzultantkou této práce.

3.2.8 Vývoj a stanovení metriky

Pro analytické zpracování dat bylo nezbytné stanovit metriku, která umožní systematické hodnocení a interpretaci získaných výsledků. Během fáze rešerše bylo identifikováno několik metodologických přístupů, které byly aplikovány v předchozích studiích na podobné téma. Tyto metody však nebyly, vzhledem ke specifikům výzkumného zaměření a použitých dat, přímo aplikovatelné v kontextu této práce.

Z tohoto důvodu bylo, po konzultaci s vedoucím práce prof. Voženílkem, rozhodnuto o vytvoření specifické metriky, která by odpovídala unikátním potřebám a cílům této práce. Diplomová práce si klade za cíl prozkoumat prostorovou podmíněnost nářečních variant a geografické distribuce vybraných příjmení v České republice, přičemž klíčovým ukazatelem je míra geografické shody těchto dvou faktorů. V počáteční fázi proto bylo navrženo vytvořit jednu metriku, která by tuto shodu umožnila ověřit a tím potvrdit či vyvrátit stanovenou hypotézu.

Po dodatečném studiu literatury a přípravě datových sad bylo zjištěno, že původně zamýšlená metrika míry shody nezohledňuje dostatečně četnost příjmení ve vztahu k počtu obyvatel nebo k celkovému počtu nositelů stejného příjmení a je spíše kvalitativní metrikou. Tento ukazatel se ovšem ukázal být jedním z klíčových aspektů pro interpretaci prostorové podmíněnosti a jeho nezohlednění by mohlo vést k významným nejasnostem ve výsledcích a snížit tak celkový přínos práce.

Výzvou při tvorbě jednotné metriky, která by zahrnovala i kvantifikovanou míru, byla rozdílná distribuce příjmení v jednotlivých ORP a velké rozdíly v celkové četnosti vybraných příjmení v České republice. To znemožňovalo standardizaci metriky na jednotnou bázi a vedlo ke zkreslení výsledků v závislosti na počtu obyvatel nebo počtu ORP, které do analýz vstupovaly.

V reakci na tyto výzvy byly nakonec vyvinuty dvě metriky, které vyjadřují jak nekvantifikovanou, tak kvantifikovanou míru shody:

- **Nekvantifikovaná míra geografické shody příjmení a nářečí (M)**
- **Index významnosti příjmení (IVP)**, který **kvantifikuje** míru geografické shody mezi příjmením a nářečím

Obě metriky byly průběžně konzultovány s PhDr. Martinou Ireinovou, Ph.D. Taková spolupráce zajišťovala, že metriky budou odpovídat odborným standardům v oblasti dialektologie.

Tento inovativní přístup umožňuje komplexnější analýzu a interpretaci dat, čímž se zvyšuje objektivita a relevance zjištěných výsledků. Metriky budou využity pro vyhodnocení dat, na základě kterých bude možné potvrdit či vyvrátit stanovenou hypotézu práce.

3.2.9 Nekvantifikovaná míra geografické shody

První definovanou metrikou je ***Nekvantifikovaná míra geografické shody příjmení a nářečí (dialekteu),*** označena písmenem *M*. Metrika na základě výsledků analýz datových sad CJA a KDE, vyhodnocuje vztah mezi dialektem a vybraným příjmením

v rámci ORP. Vztah je rozdělen do 4 kategorií, kde každá z nich reprezentuje určitý typ vztahu:

Tabulka 1 Kategorie vztahů příjmení a nářečí

kategorie	příjmení	nářeční tvar
A	výskyt	doložen
B	výskyt	**ne**doložen
C	**ne**výskyt	doložen
D	**ne**výskyt	**ne**doložen

Na základě těchto kategorií metrika ohodnocuje odpovídající ORP dle míry geografické shody. Nejsilnější míře shody odpovídá nejvyšší hodnota čísla. Hodnoty pro míru shody byly stanoveny ve spolupráci s odbornou konzultantkou následovně:

Tabulka 2 Ohodnocení kategorií dle hodnoty shody

kategorie	body	shoda
A	10 b.	100 %
B	3 b.	30 %
C	0 b.	0 %
D	10 b.	100 %

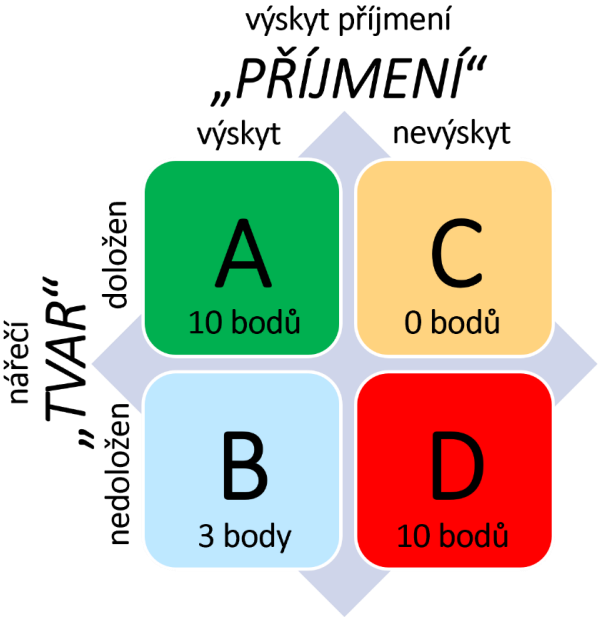
Ze stanovených kategorií vyplývá, že 100% míra shody je doložena pro ORP v kategoriích A a C (10 bodů). V těchto ORP došlo k plné shodě mezi příjmením (KdeJsme.cz) a dialektem (ČJA). Příslušná ORP, která odpovídají těmto kategoriím jsou tak ohodnocena nejvyšší hodnotou 10 bodů. V případě kategorie A zde byl potvrzen výskyt příjmení a zároveň doložen odpovídající dialekt. V případě kategorie C se v ORP příjmení nevyskytuje, ale není zde ani doložen příslušný dialekt. To znamená, že se opět jedná o shodu mezi příjmením a nářečím (shoda nevýskytu). Obě kategorie **potvrzují stanovenou hypotézu, že bydliště osob se specifickými příjmeními mají vazbu na nářeční oblasti českého jazyka.**

V případě kategorie B a C dochází k neshodě mezi výskytem příjmení a doloženým dialektem. Kategorie B v tomto případě reprezentuje oblast, ve které byl potvrzen výskyt příjmení i přesto, že zde není doložen příslušný místní dialekt. Kategorie tak jasně identifikuje oblasti, ve kterých můžeme předpokládat, že v minulosti došlo k přísunu obyvatel s určitým příjmením z oblastí, kde je výskyt takového příjmení dialektologicky doložen. Jedná se tak o regiony, které se stávají dialektologicky zajímavými pro další výzkum. Z toho důvodu není možné region hodnotit 0% shodou a byla mu stanovena míra shody 30 %.

Jedinou kategorií s 0% mírou shody tak zůstává kategorie C. V těchto oblastech se příjmení nevyskytuje i přesto, že je zde doložen příslušný dialekt. Nelze ovšem s jistotou

tvrdit, zda se jedná o region odkud došlo k pohybu obyvatel do jiných regionů nebo jde o oblast, kde se takové příjmení v minulosti nikdy nevyskytovalo. Z toho důvodu byla takovým regionům přiřazena hodnota 0.

Kategorie B a C identifikují ORP, ve kterých nedochází ke 100% shodě mezi příjmením a nářečím a proto na základě těchto samotných kategorií **nelze potvrdit hypotézu, že bydliště osob se specifickými příjmeními mají vazbu na nářeční oblasti českého jazyka.**



Obr. 5 – Kategorie nekvantifikované míry geografické shody příjmení a nářečí
(zdroj: autor práce)

Výjimka může nastat v případě, že je vybrané slovo tvořeno vlivem nářeční varianty a zároveň nářečním jevem (v mapách ohraničují izoglosy). V takovém případě může dojít ke shodě s pouze jedním z těchto vlivů. V případě tedy, že se slovo **shoduje s nářeční variantou** ale **nikoliv s nářeční jevem**, je stanovena kategorie **A2**, která je považována za 70% shodu a ohodnocena **7 body.**

Tabulka 3 Vztah příjmení a nářečí u kategorie A2

kategorie	**příjmení**	**nářeční tvar**	**nářeční jev**
A2	výskyt	doložen	**ne**doložen

Tabulka 4 Ohodnocení a míra shody kategorie A2

kategorie	**body**	**míra shody**
A2	7 b.	70 %

K tomu, aby bylo možné určit celkovou míru shody pro každé příjmení je potřeba určit celkovou **maximální hodnotu míry shody**. Jedná se o předpokládaný stav, kdy by ve všech zkoumaných oblastí (např.: ORP) došlo k maximální 100% shodě mezi příjmením a nářečím (kategorie A a C). V takovém případě by všechny oblasti nabývala hodnoty 10. Vzorec pro výpočet maximální hodnoty míry shody proto vypadá takto:

$$M_{MAX} = \text{počet zkoumaných oblastí} * 10 \quad (1)$$

Celková míra geografické shody pro jednotlivá příjmení se vypočítá určením podílu z celkové hodnoty sečtených hodnot pro všechny ORP z maximální hodnoty míry shody. Míra shody je vyjádřena v procentech. Výpočet pro určení celkové míry shody pro příjmení:

$$M_{PŘÍJMENÍ} = (\text{součet hodnot všech oblastí} / MMAX) * 100\ \% \quad (1)$$

Takovým způsobem je možné určit hodnotu M pro každé zkoumané příjmení. Tyto hodnoty lze následně porovnat mezi sebou a pomocí nich určit příjmení s nejvyšší a nejnižší mírou geografické shody. Pro vyhodnocení těchto hodnot byla ve spolupráci s vedoucím práce prof. Voženílkem a dr. Ireinovou stanovena následující škála:

Tabulka 5 Škála pro vyhodnocení celkové míry geografické shody

shoda	**míra shody**
velmi silná	více než **90 %**
silná	**66 - 90 %**
slabá	**50 - 65 %**
velmi slabá	méně než **50 %**

Tato metrika pomáhá na základě získaných procentních hodnot určit míru geografické shody pro každé zkoumané příjmení v jakékoliv stanovené oblasti výzkumu.

Pro jednodušší zapamatování pravidel metriky a pochopení vztahů mezi kategoriemi byla v průběhu řešení práce vytvořena jednoduchá pomůcka. Zde prezentována na příkladu příjmení **Žížala** (viz 3.1.1):

- kategorie **A (10 b.)** = v domě žije **pan Žížala**, na zahradě před domem leze*: **„žížala“***
- kategorie **B (3 b.)** = v domě žije **pan Žížala**, na zahradě leze: ***„červ“***
- kategorie **C (0 b.)** = v domě žije **pan Červ**, na zahradě leze: ***„žížala“***
- kategorie **D (10 b.)** = v domě žije **pan Červ**, na zahradě leze: ***„červ“***

Vizualizace v mapě je možná metodou kartogramu, pro který jsou stanoveny čtyři kategorie se striktní barevnou škálou. Pro reprezentaci „kategorií shody“ (A, D) by měly být zvoleny syté jasné barvy. Jde o oblasti, které jsou nejdůležitějším výstupem analýzy a musí v mapě vynikat. Pro zobrazení ostatních kategorií (B, C) je vhodné použít spíše bledé nevýrazné barvy, které by neměli čtenáře mapy upoutat na první pohled.

Pro zajištění správného čtení mapy by mapový list měl obsahovat jasnou legendu (např.: v podobě schématu) a jednoznačný nadpis.

3.2.10 Intenzita geografické shody

Při vyhodnocení analýzy pomocí metriky **M** (viz 3.2.9) získáme údaje o celkové míře geografické shody pro vybraná příjmení na celém zkoumaném území. Takové výsledky ovšem zahrnují pouze kvalitativní hodnocení této míry. Pro kvantifikaci těchto dat byl vytvořen **Index významnosti příjmení** označený jako **IVP**, pomocí kterého lze kvantifikovat data, která vycházejí z analýzy metrikou M. Hodnota indexu tak vyjadřuje **intenzitu geografické shody.**

Index byl vytvořen tak, aby přihlížel nejen k pouhému výskytu či nevýskytu jazykových a onomastických jevů, ale aby hodnotil intenzitu výskytu zkoumaného příjmení. Jelikož se jedná o kvantifikaci míry M, je tato metoda aplikována na ty oblasti, které byly

identifikovány mírou M jako oblasti kategorie A**.** Oblasti kategorie A2 nevykazují 100% shodu, proto na ně metrika není aplikována (viz 3.1.1). Jde o oblasti, ve kterých byla doložena nejvyšší míra shody příjmení a nářečí a zároveň jde o oblasti, které lze kvantifikovat na základě výskytu alespoň jednoho nositele příjmení. Metrika je definována dvěma parametry – **hustotou příjmení (HP) a národním poměrem příjmení (NPP):**

- hustota příjmení (HP) = vyjadřuje podíl nositelů určitého příjmení vůči celkovému počtu obyvatel v dané oblasti, přepočtený na 100 obyvatel. Je dán vzorcem:

HP = (počet nositelů příjmení v oblasti / celkový počet obyvatel v oblasti) * 100 (1)

- národní poměr příjmení (NPP) = ukazuje, jaký podíl nositelů daného příjmení v konkrétní oblasti (regionu), tvoří z celkového počtu nositelů tohoto příjmení v ČR, přepočtený na 100 obyvatel. Výpočet je proveden vzorcem:

NPP = (počet nositelů příjmení v oblasti / celkový počet nositelů příjmení v ČR) * 100 (1)

Každému parametru je ve výsledném indexu přiřazena váha (w1, w2), která vyjadřuje význam obou parametrů ve výzkumu. Pro váhy platí, že jejich výsledný součet musí dohromady udávat hodnotu 1. To znamená, že zvolené váhy musí být dosaditelné do vzorce:

$$w1 + w2 = 1 \quad (1)$$

Po testování veškerých variant nastavení vah a konzultaci výsledků s dr. Ireinovou byly váhy pro index IVP stanoveny rovnoměrně na hodnotu 0,5 pro oba parametry. Tímto způsobem je zajištěno, že oba parametry mají rovnocený vliv na konečnou hodnotu indexu. Kompletní vzorec pro výpočet indexu IVP vypadá takto:

$$\mathbf{IVP = (HP * w1) + (NPP * w2) \quad (1)}$$

kde: w1, w2 = 0,5

Výsledné hodnoty indexu značí, jak silně a významně je dané příjmení zastoupeno v jednotlivých oblastech. Tato metrika není určena k vyhodnocení hypotézy, ale pouze jako podpůrná metrika pro výsledky míry M. Na základě síly intenzity nelze určit, které příjmení má vyšší či nižší míru shody. Přidává pouze každému příjmení nebo regionu další vlastnost a spíše identifikuje regiony a příjmení vhodné pro další výzkum.

Interpretace výsledných hotdnot

Při interpretaci dat intenzity geografické shody pro regiony, jednotlivá příjmení a kompletní soubor vybraných příjmení byla použita metoda ***interkvartilového rozsahu*** **(IQR)** pro identifikaci odlehlých hodnot (outlierů) vždy v rámci jedné vybrané sady dat. Ty mohou pomoci identifikovat oblasti či příjmení s výrazně vyšší nebo nižší intenzitou geografické shody. Identifikované oblasti můžou posloužit jako podklad pro další výzkum. IQR reprezentuje rozsah hodnot, které tvoří středních 50 % proměnné a je vypočten jako rozdíl mezi třetím (Q3) a prvním kvartilem (Q1) interpretovaných hodnot (WikiScripta, 2014):

$$IQR = Q3 - Q1 \quad (1)$$

Výslednou hodnotu IQR lze pak použít ke stanovení horní a dolní hranice pro identifikaci odlehlých hodnot. Vzhledem k charakteristice vypočtených hodnot bylo přistoupeno k **vytvoření** pouze **horní hranice (HH),** která je určena součtem třetího kvartilu (Q3) a 1,5násobkem IQR (Bhandari, 2020):

$$HH = (Q3 + 1{,}5 * IQR) \quad (1)$$

Pro interpretaci dat a stanovení intenzity geografické shody pro jednotlivá příjmení byl spočítán průměr všech hodnot z vybraných oblastí. Tento průměr byl vypočten pro všechna vybraná příjmení, která vstupují do procesu kvantifikace na základě výsledků předchozí analýzy. Tyto hodnoty byly následně seřazeny tak, aby bylo možné určit a porovnat intenzitu mezi zkoumanými příjmeními. Pro vyhodnocení metriky byla stanovena následující škála, dle které lze interpretovat výsledky hodnot v rámci jednoho příjmení i pro porovnání jednotlivých příjmení mezi sebou:

Tabulka 6 Škála pro vyhodnocení intenzity geografické shody

intenzita shody	**hodnota IVP**
velmi silná	více než **HH**
silná	**Q2 - HH**
slabá	**Q1 - Q2**
velmi slabá	méně než **Q1**

V případě použití metody pro hodnocení jednotlivých regionů budou identifikovány takové, ve kterých je **koncentrován** vysoký podíl ze všech nositelů příjmení v ČR, nebo je v něm příjmení mezi obyvateli regionu častěji zastoupeno, než je tomu u ostatních regionů. Pomocí metody lze při výzkumu regionů identifikovat ty, na které je možné se zaměřit v dalším podrobnějším výzkumu. Čím silnější intenzita, tím je region významnější pro určité příjmení.

Při aplikaci této škály na hodnocení celkové intenzity pro každé příjmení, budou identifikována příjmení, které jsou koncentrovanější do regionů se 100% mírou shody a tvoří v nich vysoký podíl obyvatel. Čím **silnější intenzita**, tím je příjmení **koncentrovanější** do oblastí se 100% shodou = **regionální význam příjmení**. Naopak čím je **intenzita nižší**, tím je příjmení více **rozptýleno** po celém území = **celonárodní význam příjmení.**

Kvartily a hodnoty hranic dle IQR jsou vytvořeny z jednoho souboru dat zvlášť (hodnoty pro regiony v rámci jednoho příjmení, hodnoty pro celý soubor vybraných příjmení z jejich průměrných hodnot). Pokud by tyto hranice nebyly vypočítány zvlášť, došlo by ke zkreslení kvůli rozdílům v celkové četnosti a počtu regionů pro každé příjmení.

V mapě lze hodnoty pro oblasti v rámci jednotlivých příjmení vizualizovat metodou kartogramu, kde budou rozdílné hodnoty kvartilů znázorněny monochromatickou barevnou škálou. Hranice jednotlivých intervalů škály mohou být stanoveny výsledky statistické analýzy hodnot IVP. Jelikož jde o metriku, která je navázána na výsledky metriky M, je vhodné, aby byla použita stejná barva pro kategorii A (viz. 3.2.9) i pro hodnoty IVP v jednotlivých oblastech.

3.2.11 Vyhodnocení hypotézy

Stanovená hypotéza práce, že **bydliště osob se specifickými příjmeními mají vazbu na nářeční oblasti českého jazyka,** bude vyhodnocena na základě celkové míry geografické shody příjmení a nářečí (viz 3.2.9). Hypotézu bude možné **potvrdit** v případě, že hodnota **celkové míry geografické shody** dosáhne hodnoty minimálně **66 %.**

Hodnotu intenzity míry shody nelze samostatně použít k potvrzení či vyvrácení hypotézy práce, jelikož sama nezohledňuje prostorovou podmíněnost jevů. Slouží pouze jako podpůrná metrika, která pomáhá kvantifikovat tuto míru v oblastech s maximální mírou geografické shody. V konečné interpretaci ji lze použít k **podpoření výsledného vyhodnocení hypotézy.**

3.2.12 Analýza dat

Veškeré prostorové analýzy byly realizovány v softwaru ArcGIS Pro, který umožňuje detailní prostorové analýzy a vizualizace. Klíčovou metodou využitou při analýze byla metoda *překryvu vrstev (overlay),* která umožňuje kombinaci datových vrstev pro identifikaci vzájemných vztahů a získání nových informací z různých datových sad (ESRI, 2024c).

Data o výskytu příjmení a dialektech byla extrahována pomocí nástroje *Select by Attributes* pro identifikaci ORP s výskytem vybraných příjmení. Dále byl použit nástroj *Select by Location* pro určení ORP, kde jsou doloženy nářeční varianty vybraných slov. Nově vytvořené sloupce v atributových tabulkách, označené jako KDE pro příjmení a CJA pro dialekty sloužily jako základ pro další analýzy.

Pro kategorizaci ORP do čtyř kategorií podle stanovené metriky (viz 3.2.9) byl použit nástroj *Calculate Field,* který umožňuje zpracovávat a přepočítávat data uvnitř atributové tabulky. Tento nástroj sloužil k vytvoření jednoznačného identifikátoru pro každou kategorii spojením informací ze sloupců KDE a CJA. Tímto procesem vznikla nová datová sada **M,** která obsahuje informace nezbytné pro aplikaci stanovené metriky pro určení míry geografické shody.

Pro kvantifikaci míry geografické shody (viz 3.2.10) ve vybraných ORP, bylo nutné vytvoření čtvrté datové sady. Ta vznikla použitím nástroje *Select by Attributes*, který vybral ORP identifikované metrikou v předchozím kroku jako ORP s potvrzeným výskytem příjmení a maximální hodnotou míry geografické shody (viz 3.2.9). Tato ORP byla následně pomocí nástroje *Export Features* uložena do nově vytvořeného datasetu ***IVP*** v geodatabázi projektu. Data z těchto vrstev byla uložena do formátu .xlsx pro další zpracování.

Konečné statistické analýzy a výpočty byly realizovány v prostředí MS Excel pomocí vytvořených skriptů v programovacím jazyce Python, které automaticky načetly data a provedly potřebné výpočty. Skripty byly vytvořeny, na základě informací z veřejně dostupných internetových zdrojů (Pandas, 2024; PythonBasics, 2024; StackOverflow, 2024). Hlavní funkcí je načtení dat z vybraného Excel souboru a výpočet hodnot IVP pro každé ORP v rámci každého každé příjmení a následně vypočtení a stanovení hranic pro metodu IQR.

Pro načtení a manipulaci s daty byla využita knihovna ***pandas,*** která slouží k analýze a manipulaci s tabulkovými daty (Pandas, 2024). Ve skriptech je definována cesta ke zdrojové databázi a vzorce pro výpočty potřebné ke zjištění hodnoty požadovaných hodnot dle stanovených metrik. Pro vložení vypočtených hodnot zpět do Excel souboru je využit nástroj *Excel Writer* (StackOverflow, 2024).

Automatizace tohoto postupu výrazně zefektivnila statistické vyhodnocení zkoumaných dat v souladu se stanovenými metrikami. Postup byl ověřen průběžným testováním a manuální kontrolou, aby byla zaručena správnost výsledků. Výsledky, včetně hodnot IVP a jejich statistického vyhodnocení pomocí metody IQR, byly vizualizovány formou tabulek a map, které detailně zobrazují geografickou shodu mezi příjmeními a nářečními variantami. Oba vytvořené scripty (*vypocet_IVP.py, vypocet_IQR.py*) jsou přiloženy jako elektronické přílohy práce (viz Přílohy).

3.2.13 Technické parametry map

Veškeré mapy vytvořené pro potřeby této diplomové práce byly vytvořeny v programu ArcGIS Pro. K zobrazení administrativních jednotek byly použity příslušné vrstvy z vektorové geodatabáze ArcČR500 od společnosti ARCDATA PRAHA (2022). Mapy zobrazují území České republiky v měřítku 1: 2 893 688. Mapy jsou zobrazeny v souřadnicovém systému WGS 1984 (EPSG:4326).

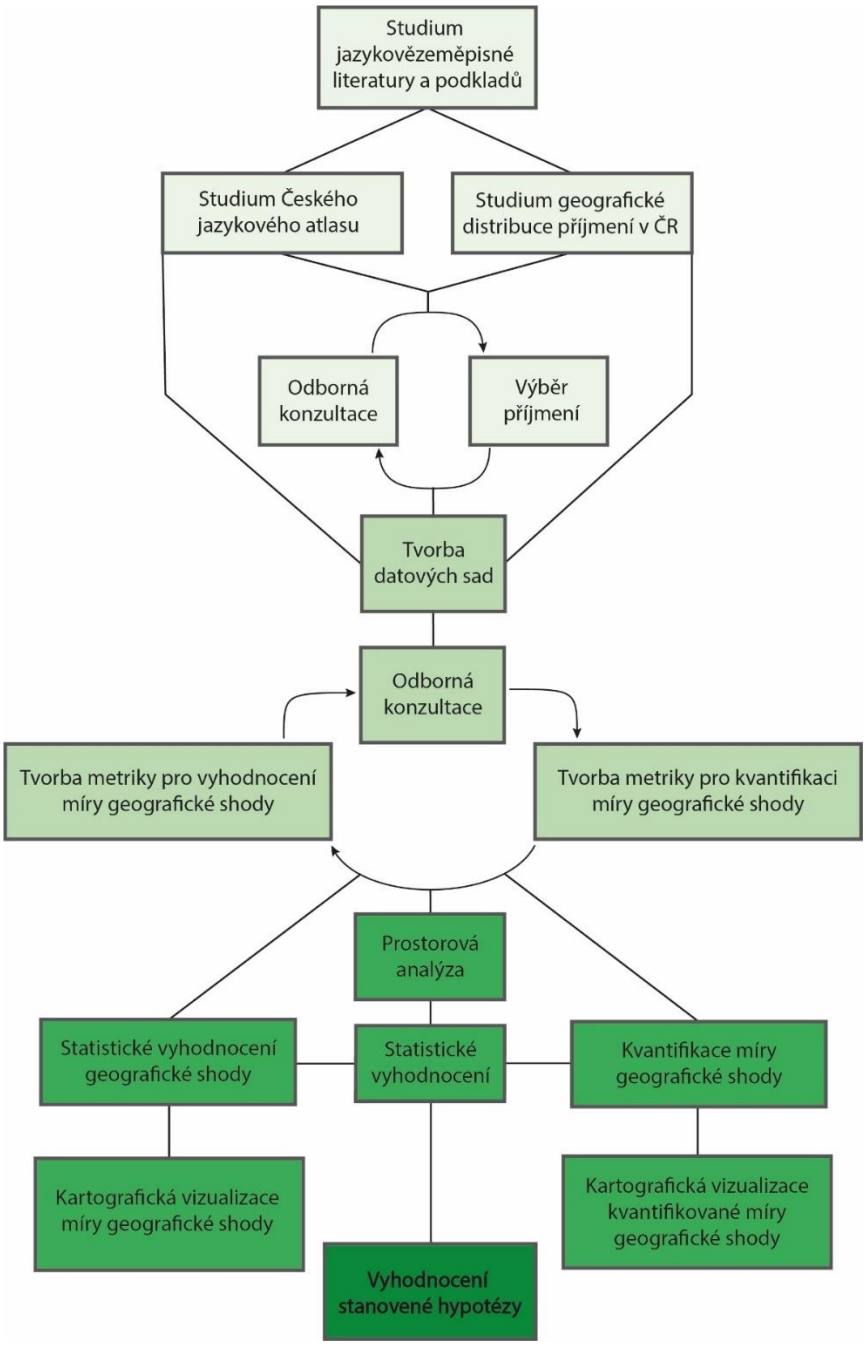
3.3 Použité programy

V diplomové práci byly využity programy, které poskytují nezbytné nástroje pro digitalizaci, georeferencování, analýzu a vizualizaci dat. Tyto programy mají zásadní význam pro zpracování a interpretaci výsledných dat.

Klíčovým programem využitím v procesu práce byl **ArcGIS Pro** ve verzi **3.2.2** od společnosti ESRI. Program umožnil vektorizaci rastrových dat z ČJA a jejich další analýzu a manipulaci. Dále byly v ArcGIS Pro provedeny prostorové analýzy využívající metod *překryvu vrstev*, které umožnily kombinování různých datových vrstev pro identifikaci vzájemných vztahů. Program byl také zásadní pro vizualizaci dat. To zahrnovalo sestavení a úpravu veškerých map, které byly v průběhu řešení práce vytvořeny.

Veškeré statistické analýzy byly provedeny v programu **Microsoft Excel**, který hrál zásadní roli při přípravě a převodu dat do prostředí GIS. Byla v něm byla statisticky zpracována data pro kvantitativní analýzu míry shody. Probíhali v něm také veškeré výpočty, které byly nezbytné pro stanovení a aplikaci metrik pro vyhodnocení výsledků. Grafy pro statistické vyhodnocení dat byly vytvořeny v prostředí webové aplikace **Fluorish.** Pro jednoduché grafické úpravy výstupů a tvorbu schémat byl použit grafický software **Adobe Illustrator CC 2019.** Veškeré využití programovacího jazyku Python probíhalo ve verzi **Python 3.11.**

3.4 Postup práce



Obr. 6 - Schéma postupu práce

Diplomová práce se zabývá analýzou geografické shody mezi příjmeními a dialekty v České republice. V práci jsou využity data z různých zdrojů, včetně Českého jazykového atlasu, webové aplikace KdeJsme.cz a geodatabáze ArcČR®500. Tato data poskytují informace o dialektech, rozložení příjmení a demografických charakteristikách ORP.

V úvodní fázi řešení práce byla studována relevantní literatura, která odpovídala zadání této práce. Studována byla tištěná literatura a použity byly také elektronické zdroje na internetu. Zkoumána byla historie dialektologického výzkumu u nás i ve světě a tvorba nářečních atlasů. Mezi taková díla patří *Český jazykový atlas* (Balhar et al., 2012) nebo *An Atlas of English Dialects* (Upton, Widdowson, 2006). Provedeno bylo také studium metodiky a výsledků studií, které analyzovaly data o příjmení v kontextu jejich geografické

distribuce ve Velké Británii a Španělsku. Ve Velké Británii se jednalo převážně o studie prováděné prof. Paulem Longleyem a jeho spoluautory. Pro pochopení základních pojmů používaných v dialektologii byly studovány příslušné odborné zdroje a práce byla konzultována s odborníky.

Ve spolupráci s odborníky bylo vybráno 15 slov a z nich vycházejících 35 příjmení. Tato data poskytla základ pro veškeré analýzy.

Ve fázi přípravy dat byla provedena vektorizace dat z ČJA. Tento proces zahrnoval převod mapových podkladů na vhodný formát, následné georeferencování a samotnou vektorizaci rastrových dat. Dále byla shromážděna data o geografické distribuci vybraných příjmení z webové aplikace KdeJsme.cz z dat MVČR z roku 2016. Napojením těchto dat na data poskytované v geodatabázi ArcČR®500 a jejich následným uložením, byly vytvořeny dvě datové sady. Ty byly následně vhodně zpracované k použití v dalších krocích.

Pro účely analýzy bylo klíčové vytvoření dvou nových metrik: Metrika pro hodnocení **míry geografické shody příjmení a nářečí** a **Index Významnosti Příjmení (IVP).** První z nich kategorizuje vztahy mezi dialekty a příjmeními na základě jejich vzájemné přítomnosti v ORP, zatímco IVP poskytuje kvantitativní hodnocení významnosti příjmení v daném regionu. Stanovení vhodných metrik je zásadním přínosem celé práce a jejich aplikace byla využita pro vyhodnocení výzkumné hypotézy práce.

Analýza dat zahrnovala prostorové a statistické analýzy, přičemž byly využity metody jako ***překryv vsrtev* (*overlay*)** a **mezikvartilový rozsah (IQR).** Výsledky těchto analýz byly vizualizovány pomocí map, grafů a tabulek, což umožnilo jasně interpretovat geografickou distribuci příjmení a jejich shodu s dialekty. K interpretaci dat byly také použity statistické metody stanovené jednotlivými metrikami. Vyhodnocení výsledků za pomoci stanovených metrik bylo klíčové pro výpovědní hodnotu celé práce.

Průběh celého výzkumu byl podrobně konzultován s odbornou konzultantkou práce **PhDr. Martinou Ireinovou, Ph.D.** z Dialektologického oddělení ÚJČ AV ČR. Tato spolupráce zajišťovala validitu metod a interpretace výsledků. Díky těmto konzultacím bylo možné průběžně testovat a upravovat výzkumné metody, což vedlo k výsledkům s vysokou mírou přesnosti a relevance pro vyhodnocení stanovené hypotézy a naplnění cílů celé práce.

Diplomová práce představuje komplexní přístup k analýze geografické shody mezi příjmeními a dialekty v České republice. Metriky stanovené v této práci jsou hlavním přínosem celé práce, jelikož byly sestaveny přímo pro účely podobného výzkumu a společně s výsledky práce poskytují nové poznatky prostorové analýzy distribuce příjmení a mohou sloužit jako základ pro další výzkum v této oblasti.

4 VÝBĚR PŘÍJMENÍ

Výběr vhodných příjmení pro analýzu byl důležitým krokem v procesu přípravy dat, klíčových pro zpracování této práce. Na základě vybraných příjmení byla vybrána také slova a odpovídající nářeční varianty, které tvoří základ vybraných příjmení. Správný výběr příjmení byl nezbytný k zajištění kvality a relevance výsledků práce. Vzhledem k potřebné expertíze, byl výběr proveden ve spolupráci s PhDr. Martinou Ireinovou, PhD., odbornicí na dialektologii z ÚJČ AV ČR.

Pro analýzu byla nakonec vybrána příjmení, která odráží jazykové rozdíly v různých částech ČR a jsou tvořena slovy, které jsou zobrazeny v ČJA. Z toho byly pro slova, tvořící základ vybraných příjmení, vybrané vhodně mapy, které byly následně použity v procesu vektorizace a poskytly základ pro tvorbu datové sady CJA. Proces výběru vhodných dat tak byl klíčový pro vytvoření obou analyzovaných datových sad. Slova a nářeční jevy zobrazené na mapách ČJA, které byly v procesu výběru příjmení určené k vektorizaci jsou:

- dešťovka
- škvor
- vrabec
- škraloup
- cop
- vrána
- mrkev
- okurka
- omáčka
- šilhavý
- vařečka
- konipas bílý
- kohout
- houser
- kukačka

Tato slova ve výběru doprovází 4 nářeční jevy, které jsou v ČJA zobrazeny na svodných mapách. Pro potřeby práce byly vybrány **svodná mapa A** a **svodná mapa D**. Geografický výskyt těchto jevů na mapě oddělují dialektické izoglosy, které oddělují území s doloženým jevem. Jedna z nich je vyznačena přímo na mapě pro slovo **omáčka.** Izoglosy oddělující jazykové jevy, které odpovídají vybraným nářečním variantám vybraných slov jsou:

Tabulka 7 Vybrané nářeční jevy ovlivňující vybraná slova

izoglosa	charakteristika	slova
A1a	krácení samohlásek	kohout / kohut
A3a	náslovné vokály (**o- X vo-**)	votápek (škraloup), okurka, omáčka
D1b	nahrazení **é, ó** X **ý, í, ú**	kohout / kohut
č. 7	**omáčka** X máčka	omáčka

Na základě zvolených slov a nářečních jevů, bylo vybráno celkem **35 příjmení** (mužská i ženská varianta příjmení), které vycházejí z různých tvarů výše zmíněných slov. a porovnávána vůči doložené nářeční variantě. Pro přehlednost bylo vytvořeno schéma, které znázorňuje vazby mezi příjmením a těmito slovy. Schéma je přiloženo jako příloha k této práci (viz Přílohy).

Důkladný výběr příjmení s vazbou na konkrétní nářeční varianty slov, představoval základ pro následné zpracování a vytvoření datových sad. Výběr tak poskytl pevný základ pro analýzy, na základě kterých byla vyhodnocována stanovená hypotéza. Spolupráce s dialektology zajistila relevanci a výpovědní hodnotu celé práce.

Jedno celé příjmení se skládá vždy z mužské i ženské varianty dohromady. V je pro přehlednost uvedena pouze jedna varianta:

- Žížala
- Škvor
- Stříhavka
- Vrabec
- Brabec
- Vrubel
- Vrábel
- Votápek
- Vrkoč
- Vrána
- Vrana
- Mrkva
- Okurka
- Vokurka
- Vokůrka
- Voharek
- Oharek
- Omáčka
- Vomáčka
- Šilhavý
- Šilha
- Šilhan
- Švidra
- Švirga
- Vařečka
- Vařecha
- Vařacha
- Vařejka
- Vařejčka
- Pliska
- Kohout
- Kohut
- Housar
- Kukačka
- Kukučka

5 GEOGRAFICKÁ DISTRIBUCE PŘÍJMENÍ

V této kapitole je představen a popsán detailní proces získávání, přípravy a zpracování dat o rozmístění zkoumaných příjmení v jednotlivých ORP České republiky.

5.1 Rozmístění příjmení v ČR

Data o rozmístění vybraných příjmení byla získána z dat Ministerstva vnitra ČR pro rok 2016, která byla poskytnuta formou webové aplikace na stránce **KdeJsme.cz** (Malačka, 2011). Pomocí vyhledávání byly zobrazeny data o distribuci mužské, i přechýlené ženské varianty příjmení. Celkem bylo vyhledáno **70 jedinečných variant příjmení**. Pro daný tvar slova byla vždy vyhledána mužská i ženská varianta příjmení zvlášť. Vyhledaná data byla ve strukturované formě uložena do vytvořené databáze ve formátu *.xlsx.*

Struktura databáze byla koncipována tak, aby přinesla co nejpřehlednější formu uložení dat a zároveň tak, aby vyhovovala požadavkům na přenos dat do prostředí ArcGIS Pro. Pro každou variantu příjmení byl vytvořen vlastní list, který nesl název daného příjmení (např.: příjmení **‚Žížala‘** → list **ŽÍŽALA**, příjmení **‚Žížalová‘** → list **ŽÍŽALOVÁ** apod.). Každý list nesl informace o četnosti příjmení v jednotlivých ORP, kde byl výskyt dle dat MVČR zaznamenán. V každém listu byl vytvořen sloupec, do kterého bylo ručně zaznamenán kód, který odpovídal kódu daného ORP dle RÚIAN (ARCDATA PRAHA, 2022). Struktura uložení dat do databáze pro každé příjmení vypadala takto:

Tabulka 8 Vhodná struktura databáze pro uložení dat

kod_ORP	**nazev_ORP**	**"PŘÍJMENÍ"_pocet**
"kód"	"název"	"počet"

Tento proces byl opakován pro každé příjmení a kompletní databáze čítá celkem 70 listů dat o geografické distribuci a počtu vybraných příjmení napříč ORP v ČR. Celkem jsou v ní uloženy tyto data pro mužské a ženské varianty 35 vybraných příjmení.

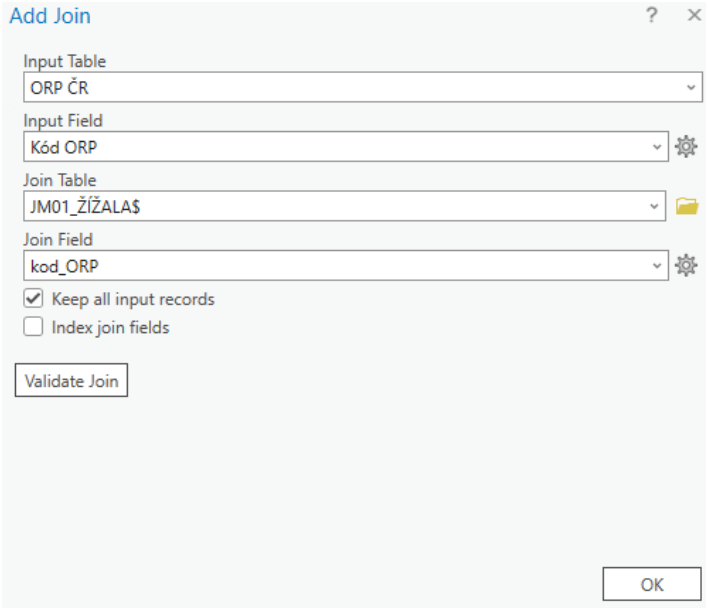
5.2 Přenos dat

V následujícím kroku bylo zapotřebí získaná data přenést do prostředí programu ArcGIS Pro, který umožní další zpracování, uložení a následnou analýzu těchto dat. Data byla uložena do databáze ve formátu, který je jednoduše umožňuje přenést do takového prostředí pomocí speciálních nástrojů určených k importu dat.

Nástroj, který umožňuje připojení dat z neprostorové databáze k vybrané vrstvě definované geografickými souřadnicemi v prostředí ArcGIS Pro je nástroj *Add Join*. Ten je dostupný jako součást nástrojů pro Data Management a na základě jedinečného identifikátoru, který musí být společný pro obě datové sady, připojuje data z jedné databáze k druhé. Data byla v tomto případě připojena k vrstvě z vektorové geodatabáze ArcČR®500 (ARCDATA PRAHA, 2022) zobrazující jednotlivá ORP v ČR.

Stejný proces byl opakován pro všechna příjmení uložené v databázi. K jedné vrstvě ORP byla vždy přiřazena data pro mužskou i ženskou variantu příjmení o stejném slovním základu (např.: Žížala + Žížalová). V práci jsou obě varianty uvažovány jako jedno příjmení. Každá vrstva byla po připojení dat pro jedno příjmení samostatně uložena do datasetu **KDE** v geodatabázi projektu. Pro uložení byl využit nástroj *Feature Class to Geodatabase.* Kompletní dataset obsahuje vrstvy, které ukládají informace o geografické distribuci a četnosti vybraných příjmení v jednotlivých ORP v ČR. Celkem bylo vytvořeno a uloženo 35 vrstev.

Tímto procesem byla získaná data o geografické distribuci vybraných příjmení přiřazena k ORP, ve kterých byl výskyt doložen. Data o četnosti jednotlivých příjmení tak byla obohacena o prostorový prvek, jelikož procesem došlo k přiřazení dat ke konkrétním geografickým souřadnicím. Na základě takto připravených dat bude možné srovnat geografickou distribuci vybraných příjmení s nářečními oblastmi ČR a umožní kartografickou vizualizaci těchto dat.



Obr. 7 - Nastavení nástroje Add Join v ArcGIS Pro (screenshot)

5.3 Zpracování dat

Pro zaručení správnosti následných analýz a vyhnutí se zbytečným problémům v procesu analýzy dat, bylo nutné data zpracovat tak, aby jejich formát odpovídal požadavkům práce a jejím cílům. Jelikož jsou mužská i ženská varianta příjmení v práci uvažovány jako jedno příjmení a data z použitého zdroje byla uvedena pouze pro jednotlivé varianty, bylo nutné tyto hodnoty sečíst.

Sečtení hodnot četnosti obou variant příjmení v daném ORP, proběhlo pro každou z vytvořených vrstev zvlášť v její atributové tabulce. V té byl nástrojem *Data Design* vytvořen nový sloupec hodnot s názvem *„PŘÍJMENÍ"_komplet*, který bude vyjadřuje celkovou četnost pro vybrané příjmení. Součet v něm proběhl pomocí nástroje *Calculate Field*, který byl nastaven tak, aby sečetl hodnoty ze sloupců, které obsahují informaci o četnosti varianty příjmení v ORP. Tento součet byl uložen do nově vytvořeného sloupce.

Během tohoto procesu byl zjištěn problém, který vznikl při připojování dat pomocí nástroje *Add Join*. Problémem byl vznik nečíselných hodnot <*Null*> ve sloupci pro četnost příjmení v jednotlivých ORP. Došlo k němu u těch ORP, u kterých nebyl výskyt dané varianty příjmení doložen – četnost příjmení byla 0. Vzhledem k tomu, že nečíselné hodnoty nemohou být sečteny, bylo nutné tyto hodnoty nahradit číselnou hodnotou 0. K tomu byl vytvořen script v jazyce Ptyhon podle návodu na stránkách technické podpory ESRI (ESRI, 2023). Kód byl uložen jako ArcGIS Notebook ve formátu. *ipynb* a spuštěn přímo v prostředí ArcGIS Pro. Spuštěn byl postupně, s odpovídajícím nastavením pro každou vrstvu.

Tímto krokem došlo k zaznamenání hodnoty 0 pro každé ORP, ve kterém nebyl potvrzen výskyt příjmení. Až po této korekci bylo možné sečíst všechny hodnoty četnosti příjmení u jednotlivých variant a získat tak kompletní data pro další kroky zpracování.

```
In [ ]:
import arcpy
path = r'C:\Users\barto\02ŠKOLA\UPOL\DIPLOMKA\DOKUMENTY\DP_prijmeni\DP_prijmeni.gdb\kde\JM01a_KDE_ŽÍŽALA'
fieldObs = arcpy.ListFields (path)
fieldNames = []
for field in fieldObs:
    fieldNames.append (field.name)
del fieldObs
fieldCount = len (fieldNames)
with arcpy.da.UpdateCursor (path, fieldNames) as curU:
    for row in curU:
        rowU = row
        for field in range (fieldCount):
            if rowU [field] == None:
                rowU [field] = 0
        curU.updateRow (rowU)
del curU
```

Obr. 8 - Nastavení kódu pro vrstvu JM01a_KDE_ŽÍŽALA (screenshot)

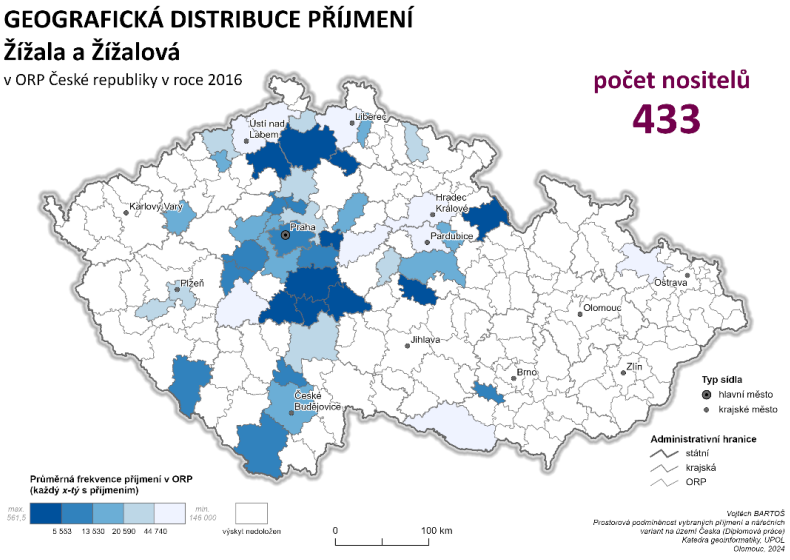
5.4 Vizualizace

Pro zjednodušení interpretace dat o geografické distribuci jednotlivých příjmení byla tato data vizualizována ve formě map. Každá z vytvořených map zobrazuje vrstvu, která odpovídá datům pro jedno příjmení. V mapě jsou zobrazeny hodnoty pro jednotlivá ORP (viz Přílohy).

Pro zobrazení byla zvolena metoda kartogramu a hodnoty použité pro stanovení kategorií reprezentují hustotu obyvatel s vybraným příjmením z celkového počtu obyvatel ORP. Tento poměr vyjadřuje, jak častý je výskyt příjmení v každém z ORP. Pro zjednodušení vyjádřen *„každý x-tý“*, kdy *x* je hodnota tohoto poměru.

Pro barevné rozlišení byla použita monochromatická barevná stupnice modré barvy. Nejtmavší barva značí nejvyšší hustotu příjmení. Kategorie hodnot byly stanoveny pomocí metody kvantilů, na základě které byl soubor hodnot rozdělen na 5 stejných dílů (pentil). Na mapovém listu je uvedena také hodnota o celkovém počtu nositelů zobrazovaného příjmení v České republice.

Pro každou vrstvu byla vyhotovena mapa dle stejných specifikací. Vytvořen byl tak soubor 35 map. Ty jsou součástí kompletního souboru map práce, který je zahrnut jako příloha (viz Přílohy).



Obr. 9 - Mapa geografické distribuce příjmení Žížala v ORP ČR
(zdroj dat: KdeJsme.cz)

Výše popsaný proces byl zaměřen na tvorbu prvního datasetu dle zadání práce. Ten obsahuje vrstvy s daty o geografické distribuci vybraných příjmení. Detailní postup a proces výběru těchto dat je popsán v předchozí kapitole. Využita byla data MVČR, která byla zobrazena ve webové aplikaci KdeJsme.cz. Za použití programu ArcGIS Pro bylo možné tato data zpracovat pro další analýzy a vhodně vizualizovat.

6 VEKTORIZACE MAP ČJA

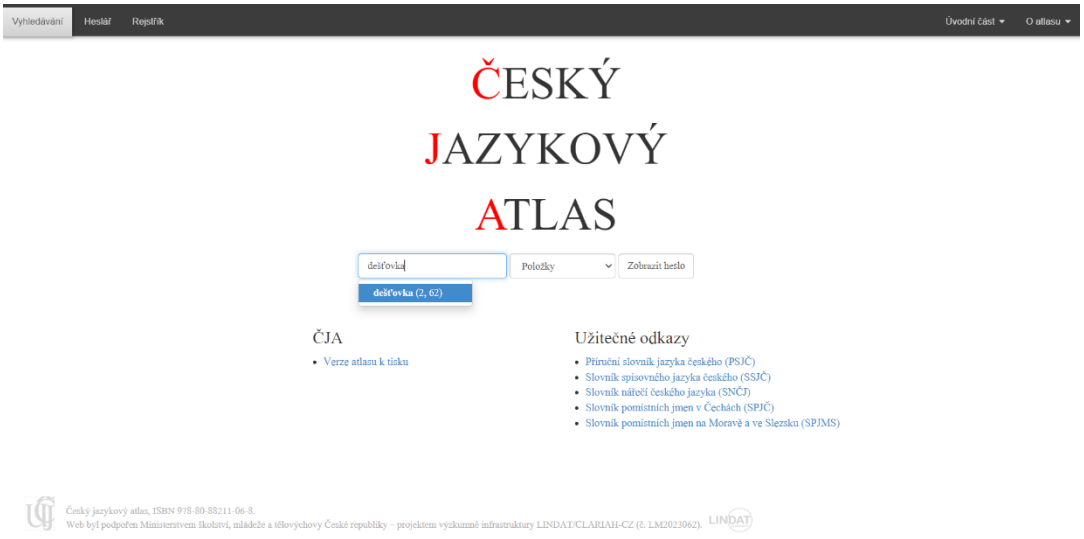
Proces transformace rastrových mapových listů z Českého jazykového atlasu byl klíčový pro tvorbu druhé sady dat, jak bylo uvedeno v zadání práce. V této kapitole bude popsán podrobný postup práce, který byl v tomto procesu použit.

6.1 Nářeční jevy na území ČR

Proces výběru slov a příjmení (viz kapitola 4) určil, u kterých je potřeba zjistit jejich rozšíření užívání v jednotlivých oblastech ČR. Tyto oblasti jsou zobrazeny v **Českém jazykovém atlase**. Ten zobrazuje dialektologická data o nářečích a nářečních variantách slov a jejich rozložení v České republice. Data, která tento atlas poskytuje, tvoří jeden ze dvou nejdůležitějších zdrojů pro vypracování této práce. Každá mapa v sobě zobrazuje informace o rozšíření užívání konkrétního slova a jeho nářečních variant. Mapy zobrazují plošné jevy, tedy oblasti, kde je daný tvar rozšířen a lokální jevy. Ty reprezentují obec nebo výzkumnou lokalitu, která se od celého území liší svým nářečím (tzv. dublety) (Hrbáček, 1974).

6.2 Výběr map

Mapy byly vybrány na základě výběru slov a příjmení, u kterých byl stanoven cíl prozkoumat jejich prostorovou podmíněnost. Proces výběru těchto slov je detailně popsán v kapitole 4. Na základě tohoto výběru byla daná slova vyhledána v elektronické podobě ČJA (Balhar et al., 2018a). Po vyhledání specifického hesla, které odpovídalo dominantnímu tvaru vybraného slova, byl zobrazen list s doprovodným komentářem o nářečních variantách slova, který byl doplněný o samotnou mapu. Výhodou použití elektronické verze atlasu bylo urychlení celého procesu o digitalizaci analogových map, která by musela v případě práce s tištěným atlasem proběhnout v přípravné fázi.



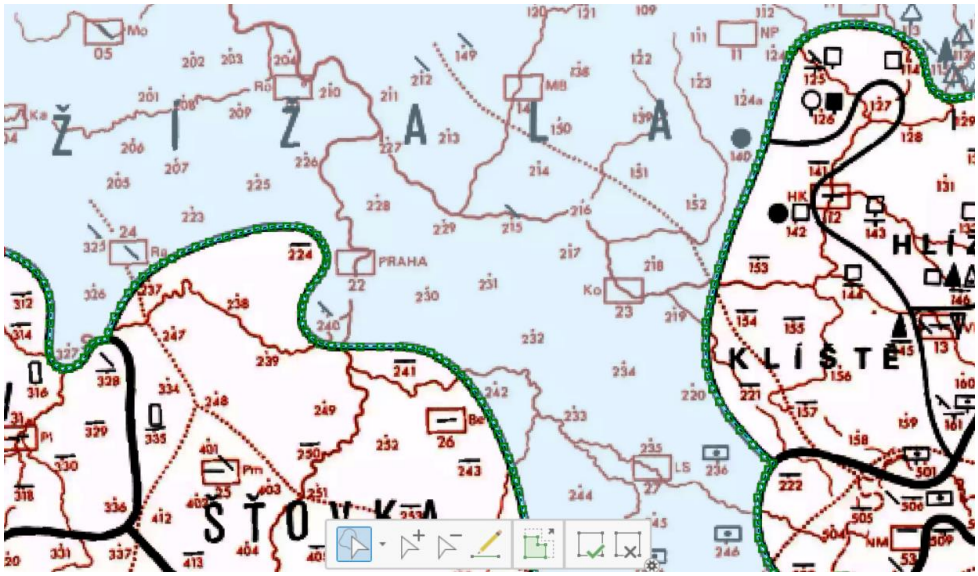
Obr. 10 - Vyhledávání v elektronické verzi Českého jazykového atlasu (screenshot)

6.3 Vektorizační proces

Při tvorbě druhé datové sady bylo nutné vybrané mapy převést z rastrového do vektorového formátu. V prvním kroku procesu byly mapy ve formátu PNG georeferencovány na vrstvu státní hranice z geodatabáze ArcČR®500 (ARCDATA PRAHA, 2022). Ke georeferenci byla

použita metoda *spline*, která požaduje vložení minimálně 10 vlícovacích bodů a poskytuje nejpřesnější výstupy (ESRI, 2024b).

Samotná **vektorizace**, byla provedena nástrojem *Create Feature* v ArcGIS Pro. Jako typ vektorizační metody byla zvolena tzv. čtvrtautomatická vektorizace využívající funkci *přychytávání bodů* (*snapping*). Vektorizovány byly všechny varianty vybraných slov, ovšem největší důraz byl kladen na slova, která přímo tvořila základ zkoumaných příjmení. Z důvodu velikosti zkoumaných administrativních jednotek, byly vektorizovány pouze **plošné jevy** (viz kapitola 3).



Obr. 11 - Proces vektorizace georeferencovaných map ČJA v ArcGIS Pro (screenshot)

Pro každou mapu byla vytvořena vlastní vrstva, která pomocí atributů rozlišovala nářeční variantu zobrazovaného slova. To následně zjednodušilo proces filtrace dat, který je potřebný pro analýzu. Pro uložení vrstev vzniklých vektorizací, byl vytvořen dataset **CJA** v geodatabázi projektu. Vytvořené vrstvy tak byly automaticky uloženy do jedné datové sady, která poskytuje kompletní informace o rozšíření vybraných slov a jejich nářečních variant, doložených v jednotlivých oblastech ČR.

Výsledné vrstvy byly na závěr oříznuté o území pohraničí, které bylo identifikováno v přípravné fázi práce. Jedná se o území převážně pohraničních oblastí, kde neprobíhal dialektologický výzkum a nejsou v něm tak doložena data o nářečí (viz kapitola 3).

6.4 Vizualizace

Vektorizované oblasti jednotlivých nářečí a nářečních variant slov, byly jako součást tvorby celého datasetu, kartograficky vizualizovány. Každá mapa zobrazuje oblast s užívanou nářeční variantou daného slova. Každá taková oblast je v rámci jedné mapy zobrazena rozdílnou barvou. Znění každé varianty je na mapě vyznačeno textem, který se nachází uvnitř odpovídající nářeční oblasti. Mapa je proto přehledná a jasně čitelná.

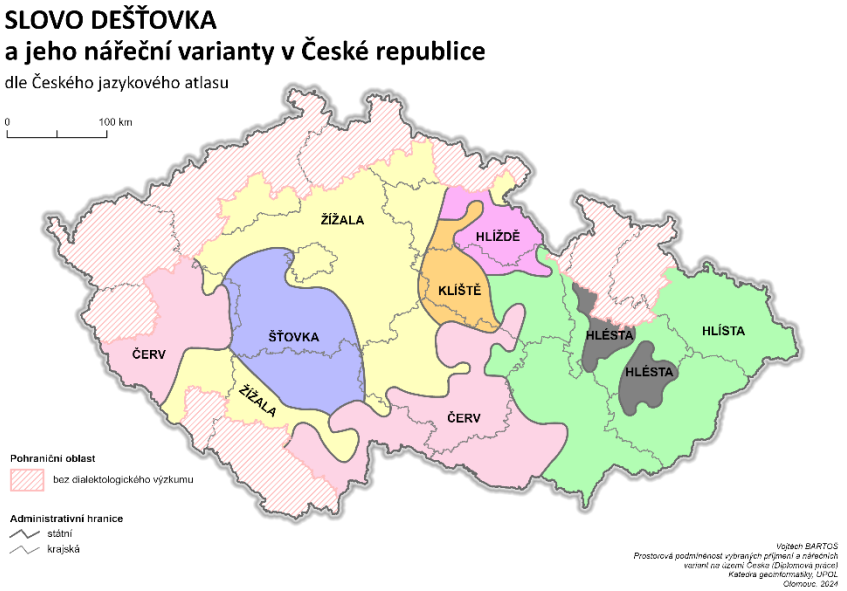
U map, které zobrazují průběh hranice vybraných izoglos, je zobrazeno celé území, ve kterém je nářeční jev užíván. Izoglosy, ohraničující toto území, jsou v mapě popsány svým označením dle ČJA.

Na mapě je dále jasně vyznačeno vymezené pohraniční území bez doloženého dialektologického výzkumu. Oblasti sousedící s pohraniční oblastí na průběhu této hranice nemají vyznačenou pevnou hranici (narozdíl jako u hranic dvou nářečních oblastí). Důvodem je stanovení pohraničních oblastí v kontextu administrativních jednotek ORP,

kterým musela být tato hranice přizpůsobena. Nejedná se proto o přesně definovanou pevnou hranici, ale pouze o hranici stanovenou pro potřeby této práce v co nejpřesnějším souladu a v závislosti na podkladových datech.

Na mapě jsou naznačeny také průběhy hranicí jednotlivých krajů pro přibližnou orientaci. Hranice ORP nejsou na této mapě naznačeny, jelikož nejsou pro téma této mapy relevantní a nepřinesly by přidanou informační hodnotu.

Celkem bylo vytvořeno 15 map zobrazujících nářeční varianty vybraných slov a 3 mapy, které zobrazují území, kde je užíván nářeční jev ohraničený izoglosou. Všechny mapy jsou jako součást souboru map práce přiloženy ve formě přílohy (viz Přílohy).



Obr. 12 – Mapa nářečních variant slova *dešťovka dle ČJA*
(zdroj dat: Český jazykový atlas)

Proces vektorizace, popsaný v této kapitole, umožnil převedení map v rastrovém formátu do vektorových vrstev navázaných na geografické souřadnice. To výrazně přispělo k efektivitě a zjednodušení následných prostorových analýz. Výsledné vektorové vrstvy, uložené v datasetu CJA, představovaly klíčový prvek pro úspěšné řešení práce.

7 EXPLORAČNÍ ANALÝZA DATOVÝCH SAD

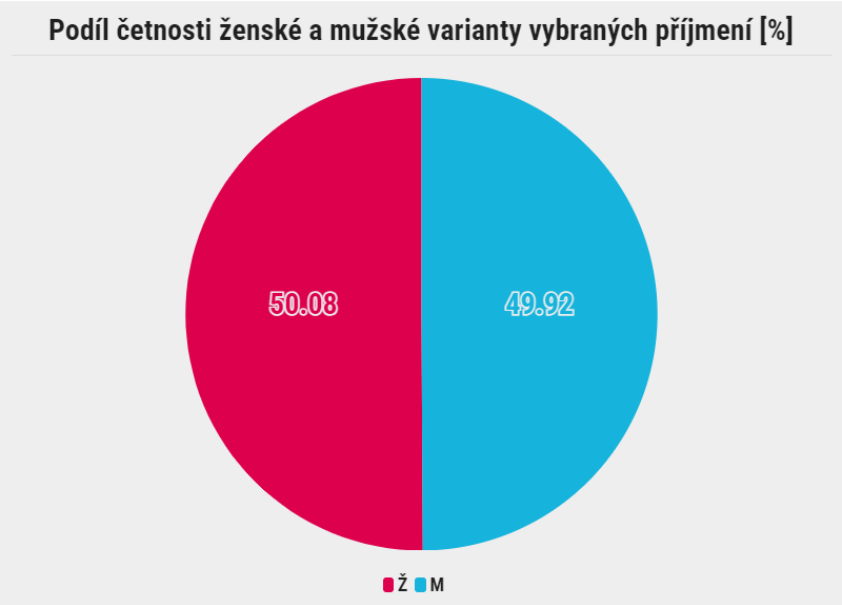
V předchozích kapitolách práce byl detailně popsán postup procesu tvorby klíčových datových sad, které byly v souladu s cíli práce vytvořeny. Vytvořené datové sady obsahují data o geografické distribuci vybraných příjmení a nářečních variantách slov, které tvoří základ těchto příjmení. Dalším krokem nezbytným k vyhodnocení stanovené hypotézy je analýza těchto datových sad. Na základě výsledků těchto analýz budou interpretovány výsledky celé práce.

7.1 Explorační statistická analýza

V první fázi procesu analýzy vytvořených datových sad byla jejich data zkoumána pomocí metod explorační analýzy. Ta pomohla pochopit základní charakteristiky vytvořených datových sad. V prvním kroku byly datové sady zkoumány pomocí nástrojů popisné statistiky. Pomocí nich byly identifikovány charakteristické hodnoty, které obě datové sady obsahují.

7.1.1 Statistické charakteristiky příjmení

Tato sada se skládá z dat o četnosti vybraných příjmení v jednotlivých ORP České republiky. Celkem datová sada poskytuje informace o četnosti a geografické distribuci **35 příjmení,** které se s různou frekvencí vyskytují na území ČR. Celkový počet nositelů všech příjmení dohromady je **36 846**. V celkovém počtu příjmení mírně převažují **ženské** varianty (**18 454 nositelů**) nad **mužskými** (**18 392 nositelů**). Vzhledem k velice malé odchylce můžeme tvrdit, že jde o datovou sadu s vyrovnaným počtem ženské i mužské varianty příjmení. Geografická shoda příjmení a nářečí tak byla zjištěna pro vzorek zhruba **0,34 %** z celkového počtu obyvatel ČR. Průměrná četnost na jedno příjmení je 1 053 nositelů.



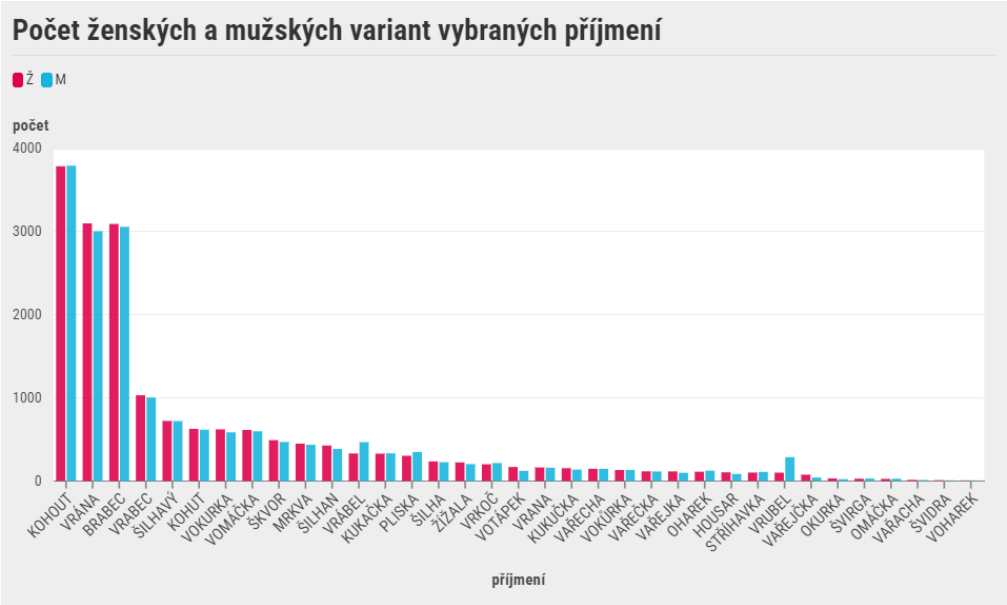
Obr. 13 - Podíl ženské a mužské varianty příjmení
(zdroj dat: KdeJsme.cz, nástroj: Fluorish Studio)

Největší absolutní rozptyl mezi počtem nositelů ženské a mužské varianty jednoho příjmení nastává u příjmení **VRUBEL**. Rozdíl mezi mužskou (291 nositelů) a ženskou variantou (102) je celkem **189** ve prospěch mužské varianty. Jediným dalším příjmením s rozdílem větším než 100 je příjmení **VRÁBEL**. Zde je rozdíl **138** shodně ve prospěch

mužské varianty (474 nositelů proti 336). Nejvyšší rozdíl ve prospěch ženské varianty příjmení je doložen u příjmení **VRÁNA**, kde je tento rozdíl **96** s převahou ženské varianty příjmení (3 144 nositelů proti 3 048).

Relativní rozdíl v četnosti, přepočtený na celkový počet nositelů je **průměrně 1 %** pro všechna příjmení. Nejvyšší je u příjmení **VRUBEL (48 %)** a **VAŘEJČKA (28 %)**. Většina příjmení ovšem dosahuje rozdílu do 10 %. Celkem tak datová sada obsahuje příjmení s nízkým rozptylem mužské a ženské varianty příjmení.

Přesná shoda četnosti obou variant nastává u příjmení **VAŘECHA**, které čítá stejný počet mužské i ženské varianty (shodně **149 nositelů**). U většiny příjmení je rozdíl v četnosti těchto dvou variant v absolutních hodnotách do 20.



Obr. 14 - Počet ženských a mužských variant příjmení
(zdroj dat: KdeJsme.cz, nástroj: Fluorish Studio)

Nejpočetnějším příjmením v datové sadě je příjmení **KOHOUT,** pod kterým je v ČR evidováno celkem **7 689 obyvatel** (KdeJsme.cz). Výrazně zastoupené jsou pak příjmení **BRABEC (6 239 nositelů)** a **VRÁNA (6 192)**. Příjmení, u kterých počet nositelů přesahuje 1 000 je v datové sadě celkem 8 a tvoří 74 % z veškerých nositelů všech vybraných příjmení.

Tabulka 9 Nejčastěji zastoupené příjmení v datové sadě (zdroj dat: KdeJsme.cz)

příjmení	počet nositelů
KOHOUT	**7 689**
BRABEC	6 239
VRÁNA	6 192
VRABEC	2 066
ŠILHAVÝ	1 464
KOHUT	1 262
VOMÁČKA	1 230
VOKURKA	1 224

V datové sadě jsou zastoupena také příjmení, u kterých je počet nositelů výrazně nižší. Jedná se o příjmení, které jsou rozšířena pouze v určitém regionu České republiky a nejsou často využívána. Příjmením s nejnižším počtem nositelů je příjmení **VOHAREK**, které v ČR nosí pouze **13 obyvatel**. Příjmení, u kterých je četnost **nižší než 50 nositelů** jsou dále **ŠVIDRA** (18 nositelů) a **VAŘACHA** (26). Příjmení s celkovým počtem nositelů nižším než 100 je v sadě celkem 6.

Tabulka 10 Nejméně zastoupené příjmení v datové sadě (zdroj dat: KdeJsme.cz)

příjmení	počet nositelů
ŠVIRGA	58
OKURKA	54
OMÁČKA	54
VAŘACHA	26
ŠVIDRA	18
VOHAREK	**13**

7.1.2 Statistické charakteristiky nářečí

Druhou datovou sadu tvoří data získaná vektorizací map nářečních variant specifických slov a nářečních jevů, které tvoří základ zkoumaných příjmení. Celkem se sada skládá z **15 map** slov a jejich nářečních variant a **3 map** zobrazujících oblasti specifických nářečních jevů. Hlavním zdrojem této sady je Český jazykový atlas.

Celkem bylo u vybraných slov zaznamenáno **76 nářečních variant**, které jsou v ČJA vyjádřeny **plošným znakem**. Slovem nejbohatším na nářeční varianty je ***škraloup*,** který má **14 variant** (např.: *škrábek, škára, svršek, otápek, kože, blaňa* atd.). **Nejnižším** počtem zaznamenaných variant u jednoho slova jsou pouze **2 nářeční varianty**. Tento počet byl zaznamenán u slov ***kohout*** (/ *kokot*)**, *omáčka*** (/ *máčka*)**, *mrkev*** *(/ mrkva).* Průměrný počet nářečních variant u vybraných slov je 5.

Obr. 15 - Počet nářečních variant vybraných slov
(zdroj dat: ČJA, nástroj: Fluorish Studio)

7.2 Explorační prostorová analýza

Jelikož jsou v datových sadách uloženy mimo statistických dat také data prostorová, bylo nutné provést základní analýzy, které pomůžou pochopit data o příjmení a nářečí v kontextu prostoru, ve kterém se vyskytují. Zjištění z těchto analýz byly základem pro další manipulaci s těmito daty. V průběhu procesu explorační analýzy byly vytvořeny také mapy zobrazující geografickou distribuci jednotlivých příjmení a rozšíření jednotlivých nářečních variant vybraných slov (viz Přílohy).

7.2.1 Prostorové charakteristiky příjmení

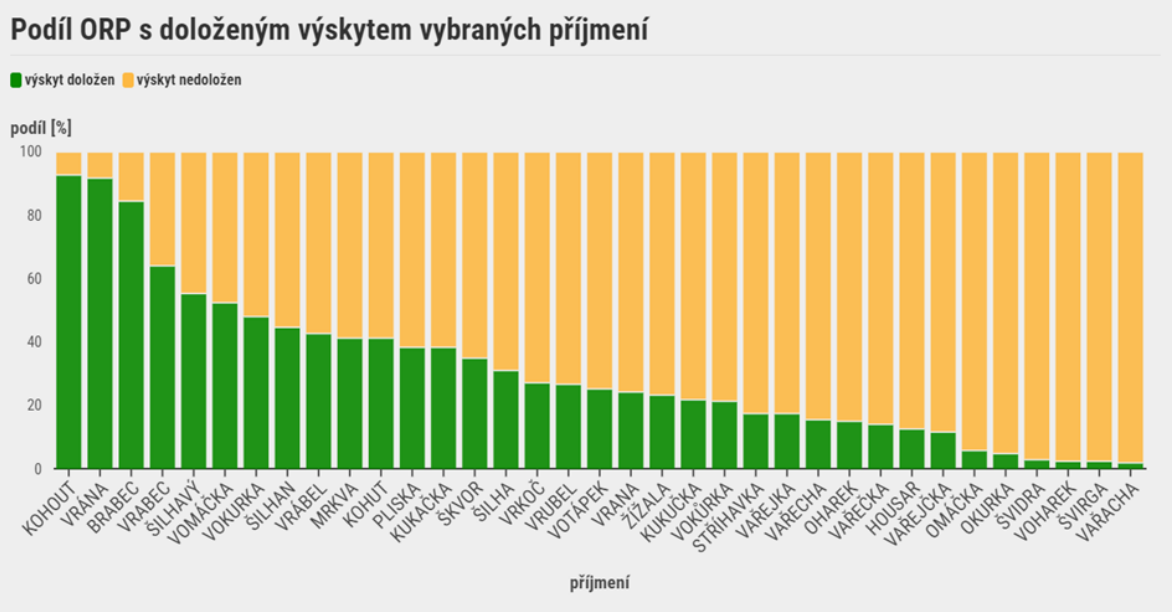
Datová sada **KDE** obsahuje informace o četnosti výskytu jednotlivých příjmení napříč všemi 206 ORP v ČR (včetně Prahy). **Nejvíce** příjmení se vyskytuje v **Praze**, kde byl doložen výskyt celkem **30 příjmení** což odpovídá více než **85 %** zkoumaných příjmení. Celkem byl výskyt více než 50 % vybraných příjmení doložen v dalších 18 ORP.

Tabulka 11 ORP s nejčastějším výskytem vybraných příjmení (zdroj dat: KdeJsme.cz)

ORP	počet příjmení
Praha	**30**
Brno	27
Černošice	24
Hradec Králové	23
Brandýs nad Labem	22

Nejméně zaznamenaných příjmení bylo shodně v ORP **Konice, Vizovice, Valašské Klobouky a Luhačovice**. V těchto regionech byly doloženy **pouze 2** z vybraných příjmení. Jedná se převážně o ORP z oblasti Valašska, které se vyznačuje specifickým nářečím. Na základě těchto poznatků lze říci, že mezi vybranými příjmeními se nacházela převážně příjmení, které nevycházejí z nářečních variant užívaných na tomto území.

Rozdílná byla mezi daty také frekvence výskytu příjmení. Zatímco některá příjmení jsou spíše regionálního charakteru a vyskytují se pouze v několika málo ORP, některá příjmení lze nalézt ve většině ORP v ČR. Nejfrekventovanějším příjmením z datové sady je příjmení **KOHOUT**, které se vyskytuje v celkem **191** ORP. Jedná se tedy o příjmení zastoupené v 93 % všech ORP. Výskyt ve více než **90 %** ORP byl doložen u příjmení **VRÁNA (189).** Nejméně frekventovaná jsou příjmení **VAŘACHA (4), VOHAREK (5), ŠVIRGA (5),** které se shodně vyskytují pouze ve **2 %** všech ORP.



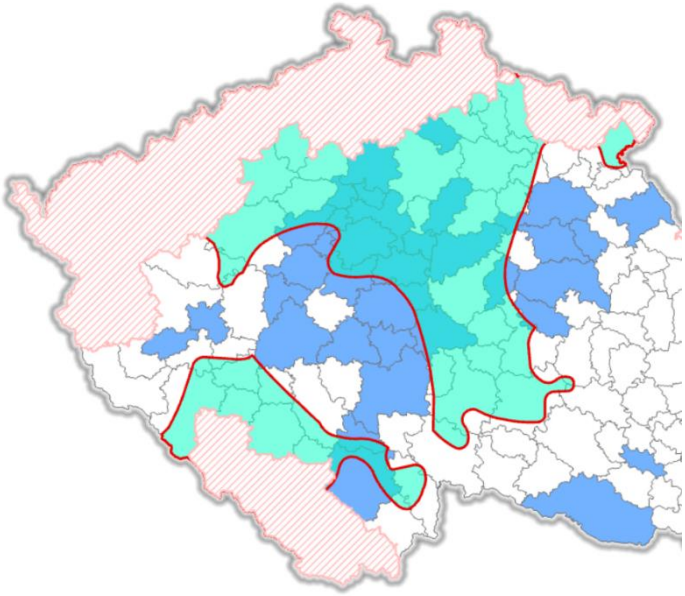
Obr. 16 – Podíl ORP s doloženým výskytem vybraných příjmení
(zdroj dat: KdeJsme.cz, ArcČR500)

7.2.2 Prostorové charakteristiky nářečí

Prvním krokem procesu analýzy vektorizovaných nářečních dat byla jejich přehledná vizualizace formou map (viz Příloha). Došlo také k analýze a vyřazení oblastí, které byly identifikovány jako pohraniční bez dialektologického výzkumu (viz 3.2.4). Následně byly

takto zpracované vrstvy vizuálně porovnány s odpovídajícími daty z datové sady geografické distribuce příjmení.

Celý proces srovnání probíhal v prostředí ArcGIS Pro, kde byly zobrazeny odpovídající vrstvy nářeční varianty a příjmení. Ty byly zobrazeny v pořadí tak, aby vrstva nářeční oblasti překrývala vrstvu o geografické distribuci příjmení v jednotlivých ORP. Následně byla vrstvě nářeční oblasti přiřazena barva výplně s průhledností 50 % tak, aby bylo možné vidět pod ní „ukrytou“ geografickou distribuci příjmení. U té byla pro lepší přehlednost změněna forma barevné vizualizace na jednotnou barvu. Při této analýze nebylo potřeba mít informace o hustotě příjmení, ale pouze o výskytu či nevýskytu.



Obr. 17 - Ukázka procesu vizuální analýzy dat v ArcGIS Pro (screenshot)

Provedená vizuální analýza dat pro všechna příjmení poskytla ucelený pohled na obě datové sady a jejich prostorovou provázanost. Díky této analýze bylo možné získat předběžnou představu a odhad prostorové podmíněnosti mezi vybranými příjmeními a nářečními variantami.

Celý proces explorační analýzy umožnil důkladné porozumění oběma vytvořeným sadám. Na základě poznatků z těchto analýz a opakovaných konzultací s dialektology bylo možné přesněji stanovit vhodné metody, které byly použity při vyhodnocování prostorové podmíněnosti příjmení a nářečí.

8 GEOGRAFICKÁ SHODA PŘÍJMENÍ A NÁŘEČÍ

Hlavním cílem práce bylo zjistit prostorovou podmíněnost mezi specifickými příjmeními a nářečními varianty vybraných slov. Připravené datové sady, které byly představeny a prozkoumány v předchozích kapitolách této práce, bylo nyní potřeba analyzovat pomocí vhodných nástrojů a metod. Tato kapitola přibližuje detailní postup a užití těchto metod ve výzkumu prostorové podmíněnosti příjmení a nářečí. Následně představí aplikaci vytvořených metrik, které byly sestaveny specificky pro tuto práci, na výsledcích provedených analýz.

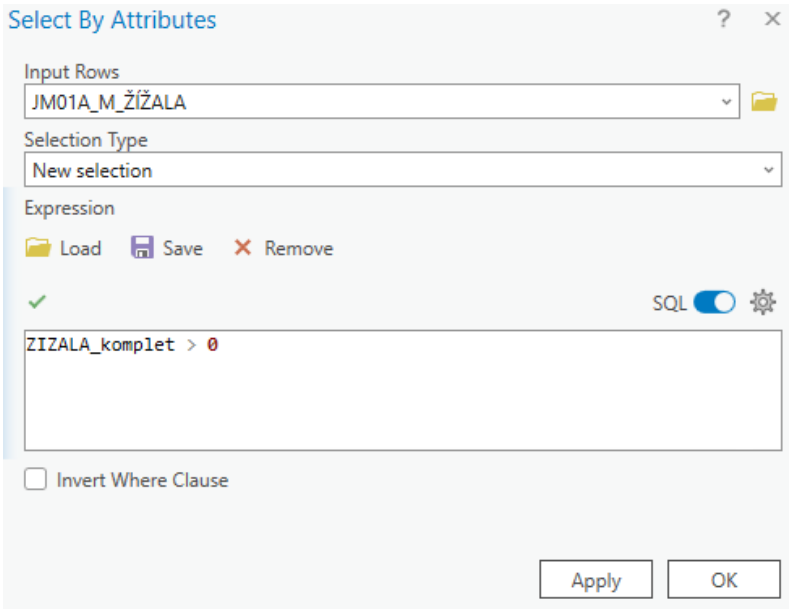
8.1 Prostorové analýzy

Připravené datové bylo nutné pomocí metod překryvné prostorové analýzy mezi sebou porovnat a na základě výsledků identifikovat oblasti, ve kterých dochází ke shodě vybraného příjmení s odpovídající nářeční variantou.

8.1.1 Identifikace ORP s výskytem příjmení

Pro účely identifikace byl vytvořen nový dataset **M** v geodatabázi projektu. Do něj byla uložena vrstva pro každé příjmení. Tato vrstva vznikla duplikací odpovídajících vrstev z datové sady **KDE** nástrojem *Export Features*. Obsahovala proto atribut četnosti daného příjmení v jednotlivých ORP. Na základě vymezení pohraničních oblastí (viz 3.2.4) byly exportovány pouze ORP, které tímto procesem byly označeny jako **oblasti s doloženým dialektologickým průzkumem.**

Pro zaznamenání informace o výskytu příjmení byl pro každou vrstvu datové sady M vytvořen v atributových tabulkách nový sloupec ***KDE***. Sloupec byl nastaven jako datový typ *Text*. Ten sloužil k zaznamenání hodnoty, která bude jasně identifikovat, zda v daném ORP došlo k potvrzení výskytu, či nikoliv. Následně byly pomocí nástroje *Select by Attributes* v každé vrstvě vybrány ORP, ve kterých byl potvrzen minimálně 1 výskyt odpovídajícího příjmení.



Obr. 18 - Natavení nástroje *Select by Attributes* v ArcGIS Pro (screenshot)

Pro všechna takto označená ORP byl do atributové tabulky a vytvořeného sloupce **KDE** zaznamenána hodnota ‚***ano***', která **potvrzovala výskyt příjmení** v ORP. Pro všechna

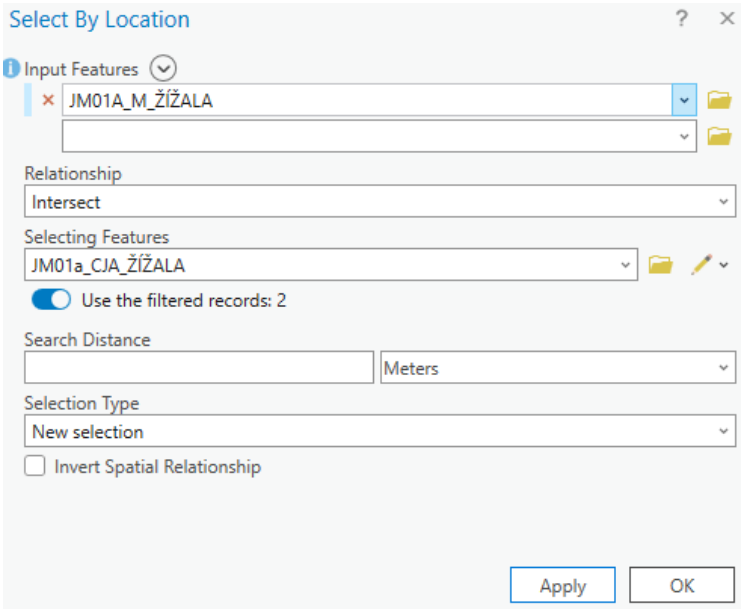
neoznačená ORP byla zaznamenána hodnota ‚***ne***', která znaší **nevýskyt příjmení** v ORP. Záznam hodnot byl proveden pomocí nástroje *Calculate Field*. Tento postup byl opakován pro všechny vrstvy vybraných příjmení.

Opakováním tohoto postupu pro všechny vrstvy vybraných příjmení bylo identifikováno **2 257 ORP**[8] s potvrzeným výskytem alespoň jednoho z vybraných příjmení. Provedeno bylo celkem 5 670 záznamů ve 35 vrstvách.

8.1.2 Identifikace ORP s doloženým nářečím

Stejný postup identifikace bylo nutné provést také pro všechna odpovídající nářečí. Jako zdroj dat sloužily vrstvy datové sady **CJA**. Identifikace jednotlivých ORP ovšem proběhla ve stejné vrstvě jako v případě identifikace výskytu příjmení.. Tedy ve vrstvě z datasetu **M** (viz. 8.1.1).

Podobně jako v předchozím kroku byl pro identifikaci vytvořen nový sloupec pro zaznamenání nářeční shody v jednotlivých ORP. Vytvořený sloupec s názvem ***CJA*** byl nastaven do datového typu *Text*. Následně byla pomocí nástroje *Select by Attributes* vyhledána odpovídající nářeční varianta z vrstvy **CJA.** Takto označená vrstva či vrstvy následně sloužily jako prvek (*Selecting Feature)*, na základě kterého byly vybrány vrstvy, které odpovídaly parametrům nastavení nástroje *Select by Location*. Ten sloužil k identifikaci ORP, které alespoň částečně protínají oblast doloženého nářečí. Jako vrstva, ze které byly na základě parametrů vybrány odpovídající prvky (*Input Feature*) byla zvolena vždy odpovídající vrstva pro jedno příjmení. Jako metoda překryvné analýzy byl zvolen *průnik* (*Intersect*).



Obr. 19 - Nastavení nástroje *Select by Location v ArcGIS Pro (screenshot)*

Pro všechny ORP označené tímto výběrem byla do atributové tabulky pomocí nástroje *Calculate Field* zaznačena hodnota ‚***ano***', která označuje **ORP s doloženou nářeční variantou** pro konkrétní příjmení. Pro ORP, kde shoda **doložena nebyla**, byla vepsána hodnota ‚***ne***'.

Při procesu identifikace těchto ORP bylo zjištěno, že je nutné také identifikovat některá sova, která dle výběru (viz kapitola 4) odpovídají nejen nářeční variantě, ale také

[8] mnoho ORP bylo identifikováno několikrát pro různá příjmení

specifickému nářečnímu jevu. Ta se v některých případech nacházela v oblasti nářečí, která by dané variantě odpovídala v případě, že by zde nebyl doložen vliv nářečního jevu. Z toho důvodu byla vytvořena další hodnota ‚***ne_izo***', která označovala ORP, kde dochází k případu, že nářeční varianta odpovídá datům z Českého jazykového atlasu, ale nářeční jev mění výslovnost tohoto slova. Příkladem je slovo ‚***omáčka***', jehož výslovnost v celém jeho území ovlivňuje nářeční jev náslovných vokálů, kdy je „***o-***„ na začátku slova nahrazeno výrazem „***vo-***„ (**o-**máčka vs. **vo-**máčka).

Postup byl opakován pro všechny vrstvy vybraných příjmení. V těchto vrstvách bylo dohromady identifikováno **2 068 ORP**, kde byl doložen výskyt vybrané nářeční varianty slova, které dalo základ jednomu z vybraných příjmení. Ve všech 35 vrstvách bylo provedeno **5 670** záznamů do atributových tabulek.

Územně nejrozšířenější nářeční variantou z celé datové sady je výraz **‚kukačka'**, který je používán téměř ve všech 162 vymezených ORP. Tento výraz se používá ve **158 ORP (98 %).** Naopak u slova **‚omáčka'** a ‚**okurka'** nedošlo ke shodě v žádném ORP. Tato slova byla ovlivněna jak nářeční variantou slova, tak nářečním jevem. V obou případech došlo k shodě pouze s nářeční variantou, ale nikoliv nářečním jevem. Ke shodě pouze s nářeční variantou došlo také v 17 ORP u slova **‚kohut',** u kterého byla ale také u 4 ORP doložena jak nářeční varianta, tak nářeční jev. Jedná se o jediné 3 případy, kdy byla použita kategorie A2 (shoda 70 %).

8.2 Aplikace vytvořených metrik

Pro potřeby práce byla ve spolupráci s odborníky na dialektologii vytvořena nová metodika pro výzkum prostorové podmíněnosti příjmení a nářečí. Tato metodika zahrnovala vytvoření metrik pro hodnocení míry geografické shody příjmení a nářečí na základě výsledků provedených prostorových analýz. Jejich vytvoření je **klíčové pro potvrzení či vyvrácení hypotézy práce.** Postup vyhodnocení vyžaduje užití dvou navzájem propojených metrik. Nejdříve je na výsledky aplikována nekvantifikovaná metrika – **míra geografické shody příjmení a nářečí (M)**. Data, která vycházejí z výsledků použití této metriky pak vstupují do procesu, který definuje metrika pro kvantifikaci míry geografické shody M – **intenzita geografické shody**, která je vyjádřena **indexem významnosti příjmení (IVP)**.

8.2.1 Míra geografické shody příjmení a nářečí

První vytvořenou metrikou byla **nekvantifikovaná míra geografické shody** (viz 3.2.9), která na základě vztahu mezi příjmením a nářečím v jednotlivých ORP definuje míru shody těchto dvou proměnných. V tomto kroku bude vytvořená metrika prakticky aplikována na datech dvou vytvořených datových sad, které byly v předchozích krocích zpracované a analyzované.

Datovou sadu **M**, která ukládá záznamy o doložení či nedoložení výskytu příjmení a nářečí v jednotlivých ORP, bylo nyní potřeba rozšířit o další hodnoty. Jako první byl vytvořen pro každou vrstvu sloupec ***M***. Na základě tohoto sloupce budou stanoveny jednotlivé kategorie a určeny hodnoty pro každé ORP. Sloupec musel být vytvořen ve stejném datovém typu jako sloupce KDE a CJA, jelikož bude obsahovat kombinaci dat z těchto sloupců. Následně byly pomocí nástroje *Calculate Field* vytvořeny 4 kombinace hodnot ze sloupců KDE a CJA, které odpovídají 4 kategoriím, které definuje metrika.

Tabulka 12 Kategorie dle vztahu mezi příjmením a nářečím a jejich ohodnocení

kategorie	KDE	CJA	M	hodnota
A	ano	ano	**anoano**	10
B	ano	ne	**anone**	3
C	ne	ano	**neano**	0
D	ne	ne	**nene**	10

Následně bylo potřeba přiřadit každé kategorii odpovídající hodnotu. Pro tu byl vytvořen v atributové tabulce každé vrstvy sloupec *hodnota.* V procesu přiřazování byl nejdříve využit nástroj *Select by Attributes,* který vybral vždy pouze ORP v jedné definované kategorii pro *M*. Pro tyto vybrané ORP byla následně pomocí nástroje *Calculate Field* přiřazena odpovídající *hodnota*.

Po přiřazení odpovídajících hodnot pro všechny kategorie, byla pomocí nástrojů ze sady *Data Engineering* zjištěna celková hodnota pro všechna ORP. Následně byla vypočtena maximální možná míra geografické shody celého souboru dat, ze které bude vypočtena míra geografické shody pro každé příjmení:

$$M_{MAX} = 162 * 10 \qquad (1)$$

$$\mathbf{M_{MAX} = 1\ 620} \qquad (2)$$

Následně byla použitím vzorce stanoveného metrikou vypočtena míra geografické shody příjmení a nářečí pro dané příjmení. Pro účely ukázky je vzorec aplikován na data pro příjmení ŽÍŽALA:

$$\mathbf{M_{ŽÍŽALA}} = (1\ 086 / 1\ 620) * 100\ \% = \mathbf{67\ \%} \qquad (1)$$

Celý tento proces byl opakován pro každé příjmení. Výsledné hodnoty byly postupně zapsány a uloženy do databáze ve formátu *.xlsx*, ve které probíhalo následné vyhodnocení těchto dat.

8.2.2 Intenzita geografické shody příjmení a nářečí

Vzhledem k charakteru zkoumaných dat, bylo nutné posoudit geografickou shody také z pohledu její intenzity. Pro tento účel byla vytvořena metrika, která kvantifikuje míru geografické shody a vyjadřuje **intenzitu geografické shody příjmení a nářečí** (viz 3.2.10)**.** Tato intenzita vyjadřuje, jak silně jsou jednotlivá příjmení zastoupena v jednotlivých ORP, kde byla odhalena 100% shoda příjmení a nářečí (kategorie A pro **M**). Pro výpočet byl vytvořen **index významnosti příjmení**, na základě kterého je intenzita stanovena. Výpočet zohledňuje počet nositelů jednotlivých příjmení v ORP, celkový počet obyvatel v ORP a celkový počet všech nositelů vybraného příjmení v rámci ČR. Index je tvořen dvěma parametry, kterým je přiřazena odpovídající váha. Tento vážený index byl vytvořen jako součást metodiky.

Výpočet indexu proběhl pro všechny ORP, které byly pomocí míry geografické shody identifikovány jako oblasti se 100% shodou příjmení a nářečí. Pro každou vrstvu tak bylo nutné pomocí nástroje *Select by Attributes* a *Export Features* vybrané ORP uložit do nového

datasetu ***A*** jako nové vrstvy. Tento dataset se tak skládal pouze z vybraných ORP, které dokládají 100% shodu příjmení a nářečí.

Následně byly vrstvy pro všechny identifikované ORP exportovány jako tabulka ve fomátu *.xlsx* z prostředí ArcGIS Pro do MS Excel. K tomu byl využit nástroj *Table to Excel*, kterým byly exportovány data ze všech vytvořených vrstev datasetu ***A***. Tímto procesem vznikl excelový soubor *prijmeni_A_ORP.xlsx.* Výpočet hodnot samotného indexu proběhl v prostředí MS Excel. Pro automatizaci výpočtu pro každé příjmení byly vytvořeny skripty v programovacím jazyce Python (*vypocet_IVP.py, vypocet_IQR.py*).

Prvním krokem byla úprava struktury databáze do požadovaného formátu tak, aby mohl být připravený skript aplikován. Každá exportovaná vrstva odpovídala jednomu listu v databázi, který byl pojmenován stejně jako samotná vrstva. Pro přehlednost byl každý z listů přejmenován podle odpovídajícího příjmení (např.: *JM01a_A_ŽÍŽALA* → *ŽÍŽALA*).Pro každý list byl zkontrolován formát sloupce tak, aby měly obdobné sloupce ve všech listech stejný název u všech listů. To zaručovalo správnost načtení a výpočtu dat pomocí připraveného skriptu. Korektní struktura databáze v každém listu vypadalo takto:

Tabulka 13 - Požadovaná struktura vstupních dat pro skript pro výpočet IVP

kod_ORP	**nazev_ORP**	**pocet_obyvatel**	**příjmení_celkem**
kód ORP	název ORP	počet obyvatel celkem v ORP	počet nositelů příjmení v ORP

Nakonec byl do databáze přidán list *příjmení ČR*, který poskytuje data o c**elkovém počtu nositelů** pro jednotlivá příjmení **v celé ČR** a list *váhy 1*, který přiřazuje váhy **w1** a **w2** potřebné k výpočtu (viz. 3.2.10). Přesná struktura a hodnoty z těchto listů jsou nutné pro správné fungování automatického výpočtu. Celá databáze byla uložena jako *prijmeni_A_ORP*.xlsx a byla načtena skriptem pro výpočet IVP *(vypocet_IVP.py*).

Hlavní funkcí vytvořeného skriptu je načtení dat ze souboru *prijmeni_A_ORP.xlsx,* který obsahuje požadovaná data o příjmeních. K tomu slouží nástroje z knihovny *pandas.* Po načtení těchto dat skript vypočítal hodnotu parametru **HP** a **NPP** pro každé ORP v jednotlivých listech (tj. pro každé příjmení). Na základě těchto hodnot vypočítal pro každé ORP ve všech listech hodnotu **IVP**. Výpočet je proveden na základě stanovených vzorců pro výpočty těchto parametrů a obecného vzorce pro výpočet **IVP** (viz 3.2.10). Ty jsou ve skriptu přesně definovány. Následně vytvořil nový list, kam uložil průměrné hodnoty IVP pro jednotlivá příjmení. Nový soubor byl uložen jako *prijmeni_IVP.xlsx.*

```
15  # Procházení jednotlivých listů a výpočet HP, NPP a IVP
16  prumery_IVP = []
17  for list in xls.sheet_names:
18      if list not in ['příjmení ČR', 'váhy 1']:  # Přeskočení shrnutí a listu s váhami
19          data = pd.read_excel(xls, sheet_name=list)
20
21          # Výpočet HP
22          data['HP'] = (data['prijmeni_celkem'] / data['pocet_obyvatel']) * 100
23
24          # Výpočet NPP
25          celkovy_pocet = celkove_pocty.get(list, 1)  # Výchozí hodnota 1 pro případ, že není nalezeno
26          data['NPP'] = (data['prijmeni_celkem'] / celkovy_pocet) * 100
27
28          # Výpočet IVP
29          data['IVP'] = 0.5 * data['HP'] + 0.5 * data['NPP']
30
31          # Zápis do nového Excel souboru
32          data[['kod_ORP', 'nazev_ORP', 'HP', 'NPP', 'IVP']].to_excel(writer, sheet_name=list, index=False)
33
```

Obr. 20 – Ukázka skriptu (.py) použitého pro výpočet IVP (screenshot)

Po vypočtení hodnot IVP pro každé příjmení bylo potřeba data vhodně interpretovat a vyhodnotit. To bylo provedeno v souladu s vytvořenou metodikou (viz 3.2.8), která stanovuje jasnou škálu pro hodnocení této metriky. Škála je sestavena pro hodnocení pomocí metody **mezikvartilového rozpětí (IQR)**, díky které je možné identifikovat odlehlé hodnoty z datových sad. Proto bylo nutné stanovit hraniční hodnoty, na základě kterých budou data vyhodnocena.

Pro tento výpočet byl opět vytvořen skript v jazyce Python, který stejně jako skript pro výpočet IVP, umožnil načíst excelový soubor. Načten byl soubor *prijmeni_IVP.xlsx*, který byl vytvořen v rámci výpočtu hodnot IVP. Ten obsahuje hodnoty IVP pro jednotlivé ORP a zároveň jejich průměrné hodnoty pro každé příjmení. Skript slouží k výpočtu hodnot hranic z těchto průměrů na základě metody IQR. Výpočtem dle zadaných vzorců IQR (viz 3.2.10) stanovil hodnoty hranic pro **1. kvartil (Q1)**, **2. kvartil (Q2)** a **horní hranici** pro identifikaci odlehlých hodnot (**HH**). Tímto výpočtem budou stanoveny hodnoty těchto hranic jak pro průměrné hodnoty v rámci jednoho příjmení, tak pro celý soubor průměrných hodnot všech příjmení dohromady. Hodnoty z výpočtu byly poté uloženy do nového Excel souboru *prijmeni_IQR.xlsx*.

```
13  # Iterace přes každý list v souboru
14  for nazev_listu in nazvy_listu:
15          data = pd.read_excel(xls, sheet_name=nazev_listu)
16          if 'IVP' in data.columns:  # Ověření, zda sloupec 'IVP' existuje
17              q1 = round(data['IVP'].quantile(0.25), 4)
18              q2 = round(data['IVP'].median(), 4)
19              q3 = round(data['IVP'].quantile(0.75), 4)
20              iqr = q3 - q1
21              horni_hranice = round(q3 + 1.5 * iqr, 4)
22              vysledek = pd.Series([nazev_listu, q1, q2, horni_hranice], index=['prijmeni', 'Q1', 'Q2', 'horni hranice'])
23              vysledky_iqr = vysledky_iqr.append(vysledek, ignore_index=True)
```

Obr. 21 – Ukázka skriptu (.py) pro výpočet hodnot hranic IQR (screenshot)

8.3 Vyhodnocení míry geografické shody

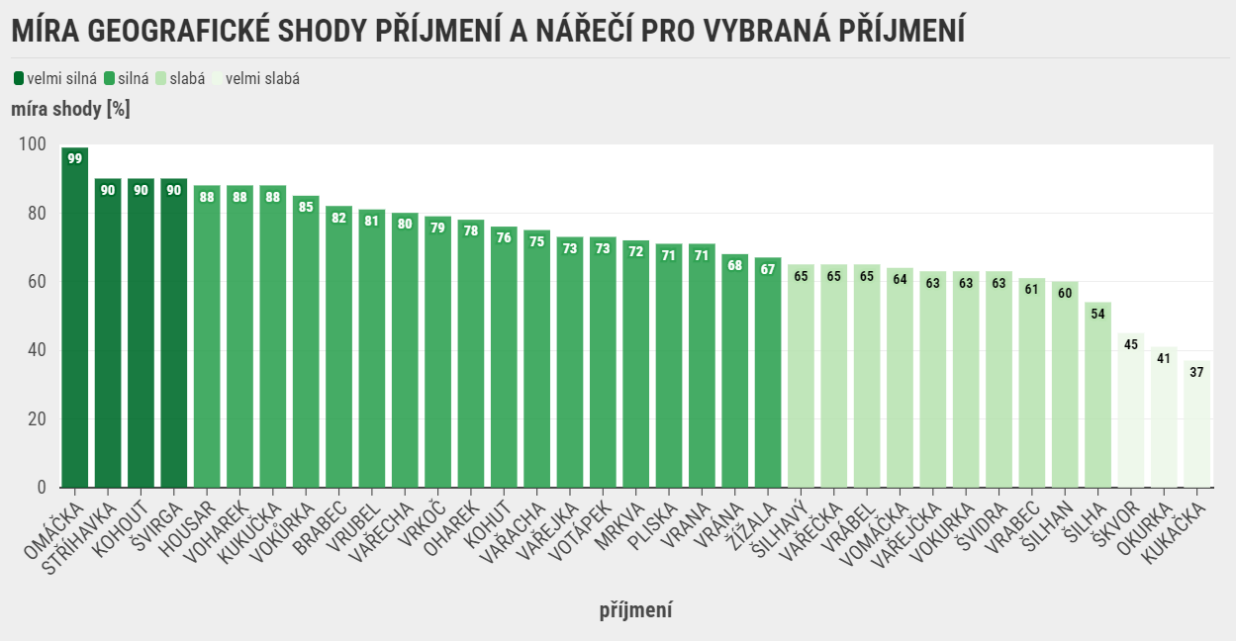
V rámci vytvořené metriky pro určení **míry geografické shody příjmení a nářečí**, byla v metodách práce definována také hodnotící škála, která slouží pro přehledné vyhodnocení výsledků a je dle ní určena celková míra geografické shody jak pro jednotlivá příjmení, tak pro celý soubor zkoumaných příjmení dohromady (viz 3.2.9).

Tabulka 14 Škála pro vyhodnocení míry geografické shody příjmení a nářečí

shoda	míra shody
velmi silná	více než **90 %**
silná	**66 - 90 %**
slabá	**50 - 65 %**
velmi slabá	méně než **50 %**

V prvním kroku byla škála použita pro vyhodnocení míry geografické shody pro jednotlivá příjmení. Nejvyšší shoda byla zjištěna u příjmení **OMÁČKA**, u kterého dosáhla **99 %**. Příjmením s nejnižší mírou shody bylo vyhodnoceno příjmení **KUKAČKA** s **37 %**. Jako příjmení s **velmi silnou shodou** byly vyhodnoceny celkem **4 příjmení**. **Silná shoda**

byla doložena pro **18 příjmení**. Slabé shody dosahovala míra u 10 příjmení a velmi slabá shoda byla určena pouze u 3 příjmení.



Obr. 22 - Vyhodnocení míry geografické shody příjmení a nářečí pro vybraná příjmení (zdroj dat: KdeJsme.cz, ČJA; nástroj: Fluorish Studio)

Na základě míry geografické shody pro jednotlivá příjmení byla vypočtena průměrná hodnota pro celý soubor což stanovilo celkovou míru shody. **Celková míra geografické shody příjmení a nářečí tak byla stanovena na 72 %.**

8.4 Vyhodnocení intenzity geografické shody

Intenzita míry shody pro jednotlivá příjmení i celková intenzita shody byla stanovena na základě hranic hodnot stanovených dle metody IQR. Pro tyto hodnoty byla stanovena škála pro rozdělení intenzity do 4 kategorií dle její síly (viz 3.2.10). Do analýzy se na základě identifikace ORP kategorie A dle míry geografické shody příjmení a nářečí, započítává celkem 33 z 35 příjmení.

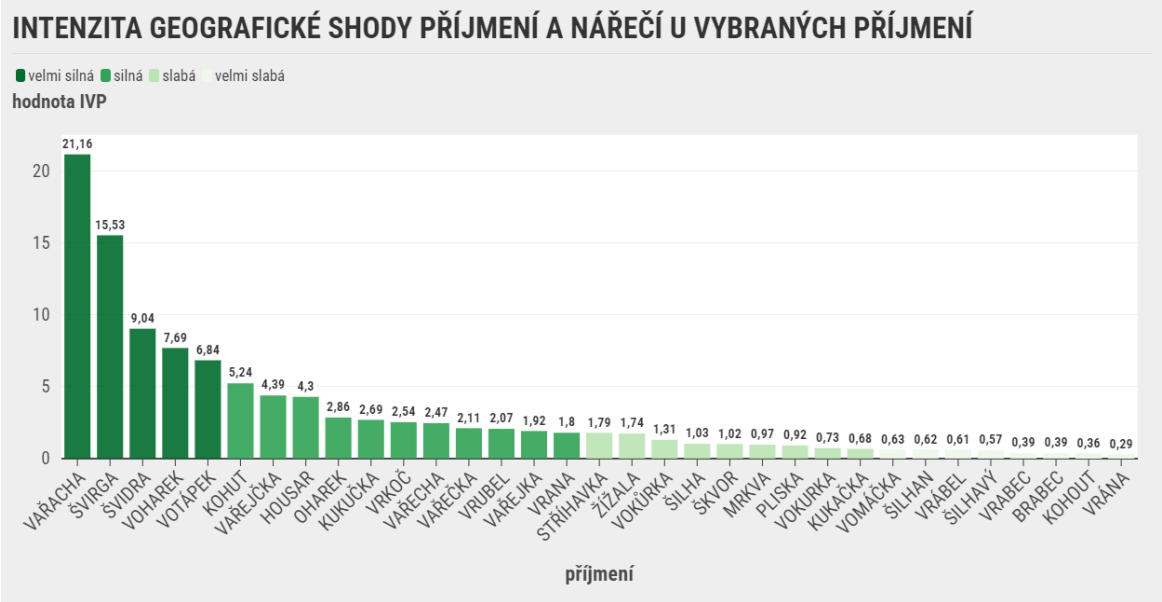
Tabulka 15 Škála pro vyhodnocení intenzity geografické shody příjmení a nářečí

intenzita shody	**hodnota IVP**
velmi silná	více než **6,1304**
silná	**1,7949 – 6,1304**
slabá	**0,6772 – 1,7948**
velmi slabá	**méně než 0,6772**

Dle hodnotící škály byla pro celkem 5 příjmení intenzita míry vyhodnocena jako **velmi silná (VAŘACHA, ŠVIRGA, ŠVIDRA, VOHAREK, VOTÁPEK)**. Hodnoty IVP pro tato příjmení byla identifikována jako výrazně vyšší, než tomu bylo u všech ostatních příjmení. Z toho lze usuzovat, že se jedná o příjmení, která jsou **silněji koncentrovaná** do ORP, ve

kterých byla prokázána **100% shoda** příjmení a nářečí. Tato příjmení lze považovat za příjmení **regionálního významu.**

Naopak výrazně nejnižší hodnoty IVP vykazovala příjmení jako **VRÁNA, KOHOUT, BRABEC** nebo **VRABEC.** Jedná se o početná příjmení a slabá intenzita značí jejich vyšší rozptýlení po celé ČR. Tato příjmení nejsou tak silně svázána pouze s určitými ORP a regiony, ale jsou oblíbená a často užívaná ve velkém území i mimo „ORP shody". Tyto příjmení jsou na základě hodnot IVP interpretována jako příjmení s **celonárodním významem.**



Obr. 23 - Vyhodnocení intenzity geografické shody příjmení a nářečí pro vybraná příjmení (zdroj dat: KdeJsme.cz, ČJA; nástroj: Fluorish Studio)

Průměrná hodnota IVP pro celý souboru vybraných příjmení je **3,23** což odpovídá **silné intenzitě shody.** Na základě výsledné hodnoty lze tvrdit, že příjmení vybraná pro tuto práci vykazují spíše **charakter regionálně významných** příjmení. Jedná se tedy převážně o příjmení, které mají tendenci shlukovat se v ORP, ve kterých byla analýzou doložena 100% míra geografické shody příjmení a nářečí.

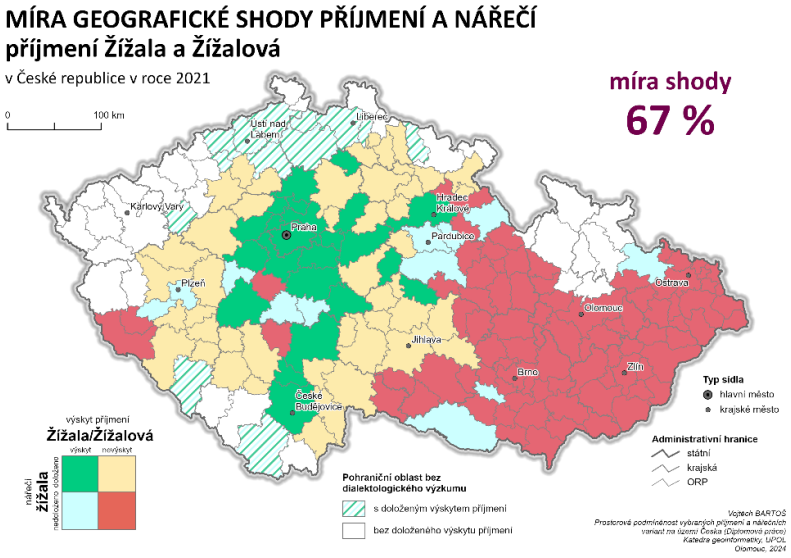
8.5 Vizualizace

Výsledky obou aplikovaných metrik byly zobrazeny v mapách. Pro každou metriku byla vytvořena jedna sada map. Mapy zobrazují ORP, krajské a státní hranice a bodovými znaky jsou zobrazeny hlavní město Praha a všechna krajská města ČR pro jednodušší orientaci při čtení mapy. Společně s ostatními mapami jsou obsaženy v souboru map, který je zahrnut v přílohách práce (viz Přílohy).

8.5.1 Mapy míry geografické shody příjmení a nářečí

Pro každé příjmení byla vytvořena samostatná mapa, která barevně rozděluje ORP na 4 kategorie dle vztahu mezi příjmením a nářečím. Pro mapu byla vytvořena přehledná schematická legenda, která přiřazuje ke každému vztahu příjmení a nářečí odpovídající barvu, kterou je zobrazeno v mapě. Vybraná ORP, ve kterých byla doložena **100% geografická shoda** mezi příjmením a nářečím jsou označeny **zelenou** (shoda výskytu) a **červenou barvou** (shoda nevýskytu) (viz 3.2.9).

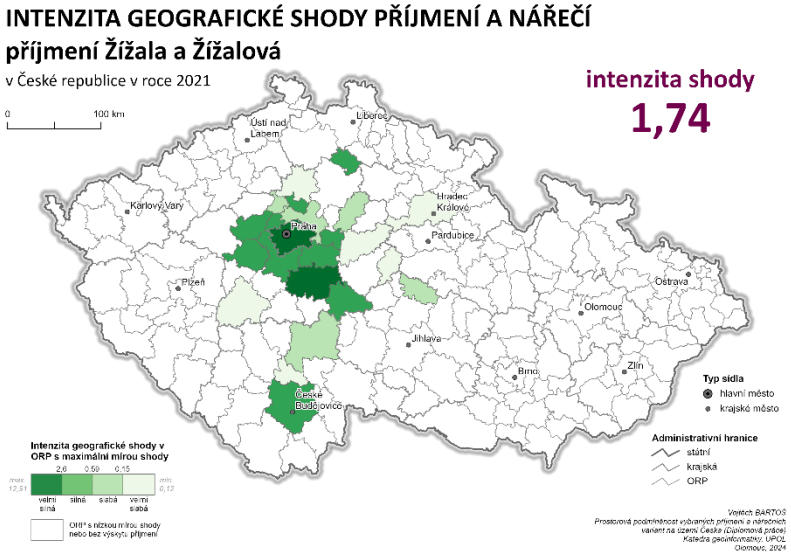
V mapě jsou dále zobrazeny také ORP v pohraničních oblastech, které byly vyřazeny z analýzy, ale byl v nich doložen výskyt příjmení. Taková ORP jsou na mapě zobrazena pomocí zeleno-modro-bílých šrafů. Na mapovém listu je také uvedena celková míra shody pro každé studované příjmení.



Obr. 24 - Mapa zobrazující míru geografické shody příjmení a nářečí v ORP ČR (zdroj dat: KdeJsme.cz, ČJA, ArcČR500)

8.5.2 Mapy intenzity geografické shody

Jelikož se jedná pouze o podpůrnou metriku, byly pro vizualizaci intenzity geografické shody vytvořeny dvě modelové mapy s návrhem možné vizualizace. Kompletně vizualizovány byly pouze sady, které se přímo podílely na vyhodnocení hypotézy. Mapa zobrazuje hodnoty IVP pro každé ORP, které jsou kategorizované dle metody IQR (viz 3.2.10). Pro barevné zobrazení byla použita monochromatická zelená stupnice. Čím tmavší barva, tím vyšší intenzita geografické shody. Vedle samotné mapy je na mapovém listě uvedena také průměrná hodnota IVP pro zobrazované příjmení.



Obr. 25 - Mapa zobrazující intenzitu geografické shody příjmení a nářečí ve vybraných ORP ČR (zdroj dat: KdeJsme.cz, ČJA, ArcČR500)

9 VÝSLEDKY

Hlavním cílem této diplomové práce bylo zjistit prostorovou podmíněnost mezi vybranými příjmeními a jejich nářečními variantami na území Česka. Tato podmíněnost byla zjištěna prostorovými analýzami 2 připravených datových sad geografické distribuce příjmení a vektorizovaných map z Českého jazykového atlasu. Pro vyhodnocení výsledků analýz byly vytvořeny 2 metriky, které umožnily vyhodnotit hypotézu, že bydliště osob se specifickými příjmeními korelují s nářečími oblastmi českého jazyka.

9.1 Vybraná příjmení

Ve spolupráci s PhDr. Martinou Ireinovou, Ph.D. z Ústavu pro jazyk český AV ČR bylo vybráno celkem 15 slov a 3 nářeční jevy, které tvoří základ 35 příjmení. Pro všechna příjmení byla prozkoumána prostorová podmíněnost jejich geografické distribuce s nářečními oblastmi ČR.

9.2 Vytvořené datové sady

Jedním z postupných cílů práce bylo vytvoření dvou datových sad, které budou sloužit jako podklad pro prostorové a statistické analýzy. První sadou dat byla geografická distribuce a četnost vybraných příjmení v jednotlivých ORP v ČR. Data pro tuto sadu byla převzata z webové aplikace KdeJsme.cz, zobrazující údaje MVČR o příjmení v roce 2016. Druhá sada dat vznikla vektorizací vybraných map z Českého jazykového atlasu. Obě sady byly vytvořeny, uloženy a vizualizovány v prostředí ArcGIS Pro.

V rámci řešení vznikly další dvě datové sady. Ty vznikly postupnými kroky, které byly provedeny k dosažení konečných výsledků práce.

9.3 Metodika

Dalším postupným cílem práce bylo stanovení vhodné metriky pro vyhodnocení prostorové podmíněnosti příjmení a nářečí. K naplnění tohoto cíle byla vytvořena kompletní metodika pro zkoumání a hodnocení prostorové podmíněnosti příjmení a nářečí. Metodika obsahuje dvě nově vytvořené metriky, určené pro stanovení **míry** a **intenzity geografické shody mezi příjmením a nářečím**. Vytvořené metriky kombinují metody prostorové a statistické analýzy. Pro použití metrik k vyhodnocení dat, byly vytvořeny speciální nástroje v jazyce Python, které využívají analytických knihoven k podrobnému zpracování dat na základě vzorců stanovených ve vytvořených metrikách. Při tvorbě metodiky byly jednotlivé kroky konzultovány s odborníky na dialektologii, aby stanovené postupy odpovídaly standardům v dialektologickém výzkumu.

První část práce je proto věnována tvorbě nástrojů a postupů pro analýzu a vyhodnocení prostorové podmíněnosti příjmení a nářečí. Kompletní postup i s nástroji je následně použit pro řešení praktické části této práce. Ve její druhé části je popsáno, jak byly nově vytvořené postupy aplikovány v analýze a jak byly stanovené metriky a vytvořené nástroje použity k vyhodnocení hypotézy práce.

9.4 Soubor map

Součástí všech provedených analýz, na základě kterých byla vyhodnocena hypotéza, byla jejich podrobná kartografická vizualizace, která umožňuje jednodušší pochopení prostorových vztahů. Celkem bylo vytvořeno 86 mapových výstupu, ze kterých byl sestaven soubor map, obsahující kompletní vizualizaci dvou zdrojových datových sad a míry

geografické shody využitou při potvrzení hypotézy. Následně obsahuje dva modelové návrhy vizualizace doplňkové metriky pro kvantifikaci míry geografické shody. Soubor je přiložen jako příloha práce (viz Přílohy).

9.5 Vyhodnocení výzkumné hypotézy

Hlavním cílem práce bylo vyhodnocení hypotézy, **že bydliště osob se specifickým příjmením, má vazbu na nářeční oblasti českého jazyka.** Tato hypotéza byla vyhodnocena za použití vytvořených metrik.

Jako hranice pro potvrzení hypotézy, byla po konzultaci s odborníky stanovena hodnota odpovídající hranici **silné míry geografické shody příjmení a nářečí**. **Hypotézu lze potvrdit** v případě, že tato hodnota míry shody pro celý soubor příjmení dosáhne **alespoň 66 %.**

Celková průměrná míra geografické shody příjmení a nářečí činní **72 %**. Na základě této hodnoty **bylo možné stanovenou hypotézu potvrdit.**

Potvrzení hypotézy podporuje také závěrečná průměrná hodnota IVP **3,23**, která značí silnou intenzitu a **koncentraci** vybraných příjmení v oblastech s doloženou shodou tohoto příjmení a nářečí.

10 DISKUZE

Hlavní ambicí této práce je prozkoumání prostorové podmíněnosti vybraných příjmení a nářečních variant na území Česka. Tento výzkum je motivován výzkumnou hypotézou o vlivu nářečních variant na bydliště osob se specifickými příjmeními. Práce je v tomto ohledu první svého druhu, jelikož podobný výzkum podobného zaměření doposud nebyl proveden u nás ani v zahraničí. Existuje mnoho studií, které využívají data o současné geografické distribuci příjmení pro studium procesů vzniku těchto příjmení nebo pro prozkoumání demografických a migračních trendů v obyvatelstvu. Data o koncentraci jednotlivých příjmení v určitém prostoru byla také využita pro prozkoumání diverzity příjmení a vyhodnocení sociologických trendů, které probíhají v jednotlivých území. Výzkum dialektů a nářečních oblastí probíhá již od konce 19. století a poskytuje cenné informace a jazykových strukturách a jejich dynamice v prostoru jednotlivých území. Chybí ovšem podrobný výzkum toho, jak spolu tyto proměnné souvisí a jak se navzájem ovlivňují a jestli je možné takový vliv spolehlivě potvrdit.

Z důvodu absence materiálů s podobnou problematikou tak nebylo možné využít již vytvořené nástroje a sestavené postupy, které by mohly být vhodně aplikovány do této práce. To představovalo výzvu, jejíž řešení bylo nutné konzultovat s odborníky na dialektologii. Na základě těchto konzultací byla pro řešení práce vytvořena jedinečná metodika, která obsahuje nástroje, metriky a postupy vhodné ke studium vztahů mezi příjmením a nářečím. Navržená metodika obsahuje metriky pro určení míry geografické shody příjmení a nářečí a její následnou kvantifikaci. K použití těchto metrik práce poskytuje nástroje a postup pro jejich implementaci do výzkumu. Hlavní přínos této práce je převážně metodologický, jelikož přináší nový pohled na komplexní analýzu vztahů mezi příjmeními a jazykovými jevy a stanovuje jasné postupy pro jejich interpretaci.

Samotná metodika je v práci následně také použita, čímž demonstruje svoji využitelnost pro výzkum jazykovězeměpsiných jevů. S její pomocí je dosaženo hlavních cílů práce, které na základě této metodiky jednoznačně potvrzují vliv nářečních jevů na prostorovou distribuci příjmení na území Česka.

Práce má proto vysoký potenciál přinést důležité poznatky pro budoucí výzkum jazykovězeměpisných jevů. Přínosem v dalším výzkumu mohou být konkrétní výsledky, které tato práce přináší pomocí aplikace vytvořených nástrojů a metod. Tyto výsledky mohou pomoci identifikovat nová témata pro další zkoumání a zároveň vymezit oblasti, ve kterých má výzkum nejvyšší potenciál přinést hodnotné poznatky. Zároveň také práce přichází s komplexním návodem a specifickými nástroji, které lze v budoucím výzkumu využít a zvýšit tím efektivitu provedeného výzkumu. Výsledky a metodika práce tak otevírá nové možnosti pro výzkum vztahů mezi příjmením a nářečím z pohledu dialektologie, demografie, historie nebo jazykové geografie. Celý postup lze využít také v obdobném výzkumu pro další a rozsáhlejší soubor vybraných příjmení.

Metody řešení práce jednoznačně ukazují sílu geoinformačních systémů a pokročilých moderních technologií ve výzkumech napříč různorodými obory. Jejich přínos v moderním výzkumu je neoddiskutovatelný a jejich široká implementace má potenciál výrazně zefektivnit budoucí výzkum. Pohled na jakákoliv data při zohlednění jejich rozmístění v prostoru vždy přináší otevření nových možností pro jejich studium, analýzu a interpretaci jejich, nejen geografických, vztahů. V dnešní době je využití moderních postupů a metod naprosto klíčovým faktorem v každém výzkumném procesu. Vhodným příkladem může být právě tato diplomová práce.

Postup práce ovšem pomohl odhalit také některé hrozby a omezení, které mohou výrazně ovlivnit budoucí výzkum. Hlavním omezením budoucího výzkumu je dostupnost

a požadovaná kvalita zdrojových dat. Hrozbou, kterou tato práce odhalila je nesourodost zdrojových dat o příjmení se statistickými daty. Z důvodu tlaku na zvýšenou ochranu osobních údajů obyvatel již dnes není možné voně získat data o aktuální struktuře a rozložení jednotlivých příjmení v ČR. V práci jsou použita data pro příjmení z roku 2016, což jsou poslední volně dostupná data, která lze pro území Česka získat. V kontextu této práce je ovšem takové „stáří" dat zanedbatelné, jelikož formování nářečních jevů a charakteristik je velice pomalým procesem, který trvá několik desítek let. Data použitá v této práci proto lze považovat za aktuální a validní. V budoucnu je ovšem možné, že s přibývajícím časem dojde k degradaci v aktuálnosti těchto dat. Řešením může být zažádání o poskytnutí dat aktuálních, které dle legislativy může sbírat Český statistický úřad. Takový proces by ovšem vyžadoval čas a kroky navíc, které mohou mít vliv na efektivitu celého výzkumu.

Na některé limitace lze také narazit v použití samotné metodiky, která je právě na kvalitě vstupních dat závislá. V případě této práce je to nesourodost v určování jednotek pro výzkum nářečí a pro ukládání informací o výskytu příjmení. Data pro příjmení jsou poskytnuta v rámci jednotlivých administrativních jednotek ORP, stanovených českou legislativou. Data o nářečních oblastech jsou vytvořena na základě výzkumných lokalit, které neodpovídají administrativnímu členění České republiky, ale byly stanoveny speciálně pro dialektologický výzkum. Při porovnání těchto dat tak není možné s absolutní jistotou určit, zda každá studovaná lokalita odpovídá přesným hranicím jednotlivých administrativních jednotek. Další faktorem je také rozdílné stáří dat, jelikož dialektologický průzkum probíhal v druhé polovině 20. století, kdy některé, aktuálně používané administrativní jednotky, neexistovaly.

V kontextu této práce je tento problém spíše potencionálním problémem, na který je nutno myslet při budoucím výzkumu. Na samotné výsledky práce má minimální vliv, jelikož hlavním cílem práce je nalezení shody mezi příjmením a nářečím pro celé území České republiky a spíše, než přesná data práce odhaluje a zkoumá určité trendy napříč celým objemným souborem vybraných dat. To přináší nutnost určité generalizace, kterou jsou vlivy tohoto problému minimalizovány a nemají zásadní vliv na výsledky ani přínos této práce.

V případě výzkumu, který se bude zaměřovat na konkrétní menší území či samostatný region by takový problém mohl ovšem způsobit značné nejasnosti a nepřesnosti v interpretaci výsledků. V takovém případě je proto vhodné zvážit přesnější metody sběru nářečních dat. Pro malé území připadá v úvahu i sběr dat přímo v terénu. Ten by ovšem vyžadoval odborný přístup a přípravu, aby byla zajištěna výpovědní hodnota sbíraných dat s ohledem na téma a cíle výzkumu. Dalším řešením může být zisk detailnějších údajů o rozmístění přímení (např. pro obce či části obcí). Zde ovšem znovu nastává výše zmiňovaný problém ochrany osobních údajů a s ním spojený náročný proces získávání takových dat.

Výsledky a metodika této práce otevírá široké možnosti jejich využití a zohlednění v dalších výzkumech. Pomocí vytvořených postupů a nástrojů lze efektivně provádět analyzovat a interpretovat vztahy mezi příjmením a nářečím v jakékoliv oblasti či regionu. Samotné výsledky analýz pak mohou pomoci při identifikaci oblastí a regionů, které mají potenciál přinést hodnotné poznatky v daném oboru. Práce také identifikuje a uznává některé limitace a problémy, spojené převážně se vstupními daty, které mohou ovlivnit výsledky dalšího výzkumu. Zároveň přináší návrhy postupů, které mohou vést k řešení těchto problémů, či alespoň k minimalizaci jejich vlivu na konečné výsledky.

Cílem autora této práce je, aby vytvořené postupy, metody a získané poznatky přispěly k efektivnějšímu průběhu budoucího výzkumu jazykovězeměpisných jevů a pomohly

v implementaci moderních technologií do těchto výzkumů. Veškeré nástroje byly specificky navrženy tak, aby poskytovaly robustní základ pro další studie a usnadňovaly použití moderních nástrojů pro hlubší porozumění zkoumaným fenoménům.

11 ZÁVĚR

Hlavním cílem této diplomové práce bylo prozkoumat prostorovou podmíněnost vybraných příjmení a nářečních variant na území Česka a vyhodnotit tak hypotézu, že bydliště osob se specifickým příjmeními má vazbu na nářeční oblasti českého jazyka. K dosažení tohoto cíle byly vytvořeny dvě sady dat, které byly vyhodnoceny na základě prostorových a statistických analýz. Vyhodnocení proběhlo s využitím nově vytvořených metrik pro určení míry a intenzity geografické shody příjmení a nářečí. Metriky jsou součástí nově vzniklé metodiky, která obsahuje kromě metrik pro vyhodnocení dat ještě specifické nástroje a postupy pro analýzu jazykovězeměpisných dat. Při tvorbě metodiky byl postup průběžně konzultován s dialektology z Ústavu pro jazyk český Akademie věd České republiky, aby byla vytvořena v souladu s odpovídajícími dialektologickými standardy.

Teoretická část práce se věnuje rešerši odborné literatury a zahraničních jazykovězeměpisných výzkumů, ve kterých byly využity nástroje geoinformačních systému pro studium vztahů mezi příjmením, nářečím a prostorem. Součástí rešerše bylo také studium Českého jazykové atlasu. Všechny tyto zdroje vedly k lepšími a podrobnějšímu pochopení jazykových vztahů a jejich dynamiky ve spojitosti s jejich prostorovou distribucí a podaly jasný základ k řešení této práce. Ve spolupráci s dialektoložkou z Ústavu pro jazyk český Akademie věd ČR, PhDr. Martinou Ireinovou, Ph.D., bylo pro výzkum vybráno 35 specifických českých příjmení a 18 slov a nářečních jevů, které tvoří základ těchto příjmení.

V první fázi řešení praktické části byly vytvořeny dvě sady dat. První sada obsahuje data o geografické distribuci vybraných příjmení na území ČR a vznikla z veřejně dostupných dat z webové aplikace KdeJsme.cz. Druhá sada dat vznikla vektorizací vybraných map z Českého jazykového atlasu, které odpovídají vybraným slovům a nářečním jevům.

V druhé fázi praktické části byla vytvořena metodika pro studium vztahů mezi geografickou distribucí příjmení a nářečními oblastmi. Tvorba metodiky se skládala ze sestavení specifických metrik pro určení míry geografické shody příjmení a nářečí a její následnou kvantifikaci. Součástí metodiky jsou také specializované nástroje pro prostorovou a statistickou analýzu a následnou interpretaci jejich výsledků. Celá metodika byla vytvořena s důrazem na její využití v dalších jazykovězeměpsiných výzkumech.

V závěrečné fázi praktické části byla metodika použita při zkoumání a vyhodnocení prostorové podmíněnosti vybraných příjmení a nářečních variant. Pomocí vytvořených nástrojů a postupů byla data z vytvořených datových sad analyzována a vyhodnocena, což umožnilo potvrdit výzkumnou hypotézu a jasně interpretovat konečné výsledky práce. Práce tak rovnou posloužila jako první případová studie pro vytvořenou metodiku.

Výsledky prostorových analýz byly zobrazeny na mapách, které byly v průběhu řešení práce vytvořeny. Z nich byl vytvořen soubor map, který je přiložen k práci ve formě přílohy. Kompletně jsou vizualizovány hlavní datové sady – 2 zdrojové sady a sada míry geografické shody, na základě které je vyhodnocena analýza. Soubor dále obsahuje návrh vizualizace doplňkové metriky pro kvantifikaci míry geografické shody. Pro interpretaci výsledků statistických analýz, byly vytvořeny přehledné grafy, které pomáhají lépe pochopit konečné výsledky výzkumu.

Výše popsanými fázemi řešení práce bylo dosaženo hlavního cíle diplomové práce. Použitím vytvořené metodiky na konkrétních datech byla detailně prozkoumána prostorová podmíněnost vybraných příjmení a nářečních variant na území Česka. Na základě výsledků tohoto výzkumu práce **potvrzuje výzkumnou hypotézu, že bydliště osob se specifickými příjmeními na území Česka, má vazbu na nářeční oblasti českého**

jazyka. Práce svými výsledky a vytvořenou metodikou otevírá nové možnosti a poskytuje cenné nástroje a poznatky k dalšímu jazykovězeměpisnému výzkumu.

POUŽITÁ LITERATURA A INFORMAČNÍ ZDROJE

ADOBE. Soubory PNG [online] [cit. 2024-04-05]. Dostupné z: https://www.adobe.com/cz/creativecloud/file-types/image/raster/png-file.html.

ARCDATA PRAHA. ArcČR® 500 verze 4.1. [geodatabáze]. Praha: ARCDATA PRAHA, 2022. Dostupné z: https://www.arcdata.cz/content/dam/distributor-share/arcdata-cz/geograficka-data/arccr/licence/arccr-4-1-popis-dat.pdf.

BALHAR, Jan; JANČÁK, Pavel. Český jazykový atlas [online]. 2., upravené vyd. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, 2018a. ISBN 978-80-88211-06 8. Dostupné z: https://cja.ujc.cas.cz/e-cja/.

BALHAR, Jan; JANČÁK, Pavel. Český jazykový atlas: O slovníku [online]. 2., upravené vyd. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, 2018b. ISBN 978 80-88211-06-8. Dostupné z: https://cja.ujc.cas.cz/e-cja/o_slovniku/.

BALHAR, Jan; JANČÁK, Pavel. Český jazykový atlas [online]. 2., elektronické, opravené, doplněné vyd. Brno: Dialektologické oddělení Ústavu pro jazyk český AV ČR, 2012. ISBN 978-80-86496-66-5. Dostupné z: https://cja.ujc.cas.cz/.

BHANDARI, Pritha. How to Find Interquartile Range (IQR) | Calculator & Examples [online]. Scribbr, 2020. Dostupné z: https://www.scribbr.com/statistics/interquartile-range/.

BOBERG, Charles; NERBONNE, John; WATT, Dominic. a kol. The Handbook of Dialectology. Hoboken: John Wiley & Sons, 2018. ISBN 9781118827598.

BŘEHOVSKÝ, Martin, JEDLIČKA, Karel. Úvod do geografických informačních systémů [online]. Katedra geomatiky, Západočeská univerzita v Plzni [cit. 2024-04-05]. Dostupné z: https://gis.zcu.cz/studium/ugi/e-skripta/ugi_k3b-cinnosti_v_GIS.pdf.

CHESHIRE, James a LONGLEY, Paul. Identifying spatial concentrations of surnames [online]. *Journal of Geographical Information* Science. 2011a, s. 309-325. Dostupné z: https://doi.org/10.1080/13658816.2011.591291.

CHESHIRE, James a LONGLEY, Paul. Spatial concentrations of surnames in Great Britain [online]. *Procedia Social and Behavioral Sciences*. 2011b, 21, s. 279 – 286. Dostupné z: https://doi.org/10.1016/j.sbspro.2011.07.047.

ČESKÝ STATISTICKÝ ÚŘAD. Zákon č. 89/1995 Sb., o státní statistické službě [online]. ČSÚ, 2024. Dostupné z: https://www.czso.cz/csu/czso/zakon_o_statni_statisticke_sluzbe.

ESRI. Georeferencing Definition [online]. Technical Support, Gis Dictionary [cit. 2024a-04-05]. Dostupné z: https://support.esri.com/en-us/gis-dictionary/georeferencing.

ESRI. Overview of georeferencing [online]. Esri, ArcGIS Pro [cit. 2024b-04-20]. Dostupné z: https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/overview-of-georeferencing.htm.

ESRI. Understanding overlay analysis [online]. [cit. 2024c-04-05]. Dostupné z: https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/understanding-overlay-analysis.htm.

ESRI. Vectorization Definition [online]. Technical Support, GIS Dictionary [cit. 2024d-04-18]. Dostupné z: https://support.esri.com/en-us/gis-dictionary/vectorization.

ESRI. How to Replace null values with zeroes in an attribute table in ArcGIS Pro [online]. Technical Support 2023. Dostupné z: https://support.esri.com/en-us/knowledge-base/how-to-replace-null-values-with-zeroes-in-an-attribute-000023190.

ESRI. Understanding Raster Georeferencing [online]. 2018. Dostupné z: https://www.esri.com/about/newsroom/wp-content/uploads/2018/07/Understanding-Raster-Georeferencing.pdf.

HRBÁČEK, Josef. Lexikální ekvivalenty [online]. Naše řeč. 1974, 57, 1, s.28-33. Dostupné z: http://nase-rec.ujc.cas.cz/archiv.php?art=5742.

JUSTOŇ, Zdeněk. *Endogamie*. Encyklopedie sociologie. Praha: Sociologický ústav AV ČR. [online]. 2017a [cit. 2024-04-25]. Dostupné z: https://encyklopedie.soc.cas.cz/w/Endogamie.

JUSTOŇ, Zdeněk. *Exogamie.* Encyklopedie sociologie. Praha: Sociologický ústav AV ČR. [online]. 2017b [cit. 2024-04-25]. Dostupné z: https://encyklopedie.soc.cas.cz/w/Exogamie.

KLOFEROVÁ, Stanislava. *Dialektologie*. CzechEncy: Nový encyklopedický slovník češtiny. online]. 2017a [cit. 2024-04-24]. Dostupné z: https://www.czechency.org/slovnik/DIALEKTOLOGIE.

KLOFEROVÁ, Stanislava. *Jazykový atlas*. CzechEncy: Nový encyklopedický slovník češtiny. online]. 2017b [cit. 2024-04-24]. Dostupné z: https://www.czechency.org/slovnik/JAZYKOVÝ ATLAS.

KLOFEROVÁ, Stanislava. *Jazykový zeměpis*. CzechEncy: Nový encyklopedický slovník češtiny. [online]. 2017c [cit. 2024-04-24]. Dostupné z: https://www.czechency.org/slovnik/JAZYKOV%C3%9D%20ZEM%C4%9APIS.

KLOFEROVÁ, Stanislava. *Izoglosa*. CzechEncy: Nový encyklopedický slovník češtiny. [online]. 2017e [cit. 2024-04-24]. Dostupné z: https://www.czechency.org/slovnik/IZOGLOSA.

LONGLEY, Paul, WEBBER, Richard, LLOYD, Daryl. The quantitative analysis of family names: historic migration and the present day neighborhood structure of Middlesbrough, United Kingdom [online]. *Annals of the Association of American Geographers*. 2007, 97, 1, s. 31-48. Dostupné z: https://doi.org/10.1111/j.1467-8306.2007.00522.x.

MALAČKA, Ondřej. KdeJsme.cz. [online]. 2011. Dostupné z: https://www.kdejsme.cz/.

MATEOS, Pablo a TUCKER, Ken. Forenames and Surnames in Spain in 2004. [online]. *Names: A Journal of Onomastics*. 2008, 56, 3, s.165–184. Dostupné z: https://doi.org/10.1179/175622708X332860.

MINISTERSTVO VNITRA ČESKÉ REPUBLIKY. Co je GDPR [online]. MVČR [cit. 2024-02-25]. Dostupné z: https://www.mvcr.cz/gdpr/clanek/co-je-gdpr.aspx.

MOLDANOVÁ, Dobrava. Naše příjmení. 2. vyd. Praha: Agentura Pankrác, 2004. ISBN 80-86781-03-8.

MCELDUFF, Fiona, MATEOS Pablo, WADE, Angie, CORTINA-BORJA, Mario. What's in a Name? The Frequency of Geographic Distributions of UK Surnames [online]. *Significance.* 2008, 5, 4, s. 189–192. Dostupné z: https://doi.org/10.1111/j.1740-9713.2008.00332.x.

Pandas. API reference - pandas.read_excel [online]. NumFOCUS, Inc. [cit. 2024-04-29]. Dostupné z: https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.read_excel.html.

PLESKALOVÁ, Jana. *Onomický objekt*. CzechEncy: Nový encyklopedický slovník češtiny. online]. 2017a [cit. 2024-04-24]. Dostupné z: https://www.czechency.org/slovnik/ONYMICK%C3%9D%20OBJEKT.

PLESKALOVÁ, Jana. *Onomastika*. CzechEncy: Nový encyklopedický slovník češtiny. online]. 2017b [cit. 2024-04-24]. Dostupné z: https://www.czechency.org/slovnik/ONOMASTIKA.

Příručka ČNK, Zipfovy zákony [online]. [cit. 2024-04-24]. Příručka ČNK, 2013. Dostupné z: https://wiki.korpus.cz/doku.php?id=pojmy:zipf&rev=1379083949.

PythonBasics. Read Excel with Python Pandas [online]. [cit. 2024-04-29]. Dostupné z: https://pythonbasics.org/read-excel/.

StackOverflow. importing an excel file to python [online]. [cit. 2024-04-29]. Dostupné z: https://stackoverflow.com/questions/43964513/importing-an-excel-file-to-python.

TechSmith, JPG vs. PNG: Which is Better? [online]. [cit. 2024-04-25]. TechSmith. Dostupné z: https://www.techsmith.com/blog/jpg-vs-png/.

UPTON, Clive a WIDDOWSON, J.D.A. An Atlas of English Dialects. 2. vyd. London a New York: Routledge, 2006. ISBN 978-0-415-39233-4.

VAN DIJK, Justin a LONGLEY, Paul A. Interactive display of surnames distributions in historic and contemporary Great Britain [online]. *Journal of Maps*. 2020a, 16, 1, s. 68-76. Dostupné z: https://doi.org/10.1080/17445647.2020.1746418.

VAN DIJK, Justin a LONGLEY, Paul A. Platial Geo-Temporal Demographics Using Family Names [online]. In MOCNIK, Franz-Benjamin a WESTERHOLT René. *Proceedings of the 2nd International Symposium on Platial Information Science (PLATIAL'19).* Coventry, UK. University of Warwick, 2020b. s. 23–31. Dostupné z: https://doi.org/10.5281/zenodo.3628863.

VAN DIJK, Justin, LANSLEY, Guy, LAN, Tian, LONGLEY Paul A. Using the spatial analysis of family names to gain insight into demographic change [online]. In ROBSON, Craig. Proceedings of the 27th Conference on GIS Research UK (GISRUK). Newcastle, UK. Newcastle University, 2019. Dostupné z: https://www.researchgate.net/publication/333650494_Using_the_spatial_analysis_of_family_names_to_gain_insight_into_demographic_change.

VOŽENÍLEK, V.; KAŇOK, J. a kol. Metody tematické kartografie: vizualizace prostorových jevů. 1. vyd. Olomouc: Univerzita Palackého v Olomouci pro katedru geoinformatiky, 2011. ISBN 978-80-244-2790-4.

Wikipedia, Inkvizice [online]. [cit. 2024-04-24]. Wikipedia. Dostupné z: https://cs.wikipedia.org/wiki/Inkvizice.

WikiSkripta. Míra variability [online]. WikiSkripta, 2014. Dostupné z: https://www.wikiskripta.eu/w/M%c3%adry_variability.

Zákon č. 314/2002 Sb., o stanovení obcí s pověřeným obecním úřadem a stanovení obcí s rozšířenou působností. In: *Zákony pro lidi* [online]. AION CS, 2010–2024 [cit. 2024-04-05]. Dostupné z: https://www.zakonyprolidi.cz/cs/2002-314.

PŘÍLOHY

SEZNAM PŘÍLOH

Volné přílohy

Příloha 1 Poster

Příloha 2 Soubor map (ukázka)

Příloha 3 Vybraná příjmení

Příloha 4 Skripty

- 01_vypocet_IVP.py
- 02_vypocet_IQR.py

Elektronické přílohy

Příloha 5 Digi medium

Popis struktury odevzdávaných digitálních dat na datové úložiště katedry (Digi medium)

- Poster
- Text_Prace
- Vstupni_Data
  - Databaze
  - Mapy
- Vystupni_Data
  - Databaze
  - Skripty
  - Soubor map (komplet)
- WEB