

**Bronislav Stupňánek**

Ústav pro jazyk český AV ČR, v. v. i.

Brno

ORCID: 0000-0002-9926-2312; e-mail: stupnanek@ujc.cas.cz

## Metodologie interaktivního mapového zpracování systematických gramatických struktur\*

**Abstrakt:** Připravovaný nářeční geoportál DiaMa představí řadu metodologických inovací při zpracování nářeční gramatiky. Mimo materiál z *Českého jazykového atlasu* využije i jeho metodu systematické abstrakce prostřednictvím extrakce zejména fonologických proměnných z proměnných morfologických a lexikálních. Díky tomuto rozkladu materiálu na více atomických jevů je možné široce využít geo-statistickou techniku interpolace prostorových dat, která nejenže modeluje průběhy izoglos lépe na základě stejných dat, ale také umožňuje tato systematická dotazníková data rozšiřovat a zpřesňovat pomocí dat nesystematických, náhodně rozptýlených. Rozvržení interpolované prostorové databáze pak umožní volitelný řez mnohazměrnou gramatickou strukturou a pokročilou mapovou vizualizaci uživatelsky definovaných datových sestav.

**Klíčová slova:** dialektologická mapa, syntetická mapa, izoglosa, interpolace, bodová data, scattered data, dialektometrie, *Český jazykový atlas*.

**Abstract: The Methodology of Interactive Map Processing of Systematic Grammatical Structures.** DiaMa, an upcoming dialect geoportal, will present a number of methodological innovations in the processing of dialectal grammar. In addition to the material from the *Český jazykový atlas* [Czech Linguistic Atlas], it will also use methods of systematic abstraction by extracting mainly phonological variables from morphological and lexical ones. As a result of breaking down the material into multiple atomic phenomena, it is possible to use geostatistical techniques of spatial data interpolation on a wide scale.

These techniques model the isoglosses more exactly than a classic linguistic atlas map based on the same data. What is more, they also allow to enhance and refine systematic questionnaire data with a random set of “scattered data”, making maps of the same phenomena much more accurate. Consequently, the layout of the interpolated spatial database allows customized cross-sectional views of a multi-dimensional grammatical structure and advanced map visualisation of user-defined data sets.

**Keywords:** dialect map, synthetic map, isogloss, interpolation, point data, scattered data, dialectometry, Czech Linguistic Atlas.

---

\* Tento příspěvek vznikl na základě řešení projektu č. DG20P02OVV029 *Nářečí českého jazyka interaktivně. Dokumentace a zpřístupnění mizejícího jazykového dědictví jako nedílné součásti regionálních identit*; poskytovatel podpory Ministerstvo kultury, program NAKI II. Práce používá také data, která poskytuje výzkumná infrastruktura LINDAT/CLARIAH-CZ (<https://lindat.cz>) podporovaná Ministerstvem školství, mládeže a tělovýchovy České republiky (projekt č. LM2018101).

Cílem tematické mapy je zpravidla přehledné zobrazení nějakého prostorově rozloženého jevu. Faktor přehlednosti a čitelnosti mapy je bezpochyby důležitý – i v našem oboru známe řadu případů, kdy mapa je přetížená symboly a informacemi a pro lidské oko už téměř nic bezprostředně zjevného nezobrazuje; nechte se o nic lépe než tabulka nebo slovní popis. Jakkoli složitost může působit odborně a vědecky<sup>1</sup>, v oboru tematické kartografie je považována za nedostatek a je snaha ji eliminovat např. metodou eye trackingu<sup>2</sup>.

Nejlépe se na lingvistických mapách zobrazují jevy jednoduché (atomické) a jevy, které jsou dobře kvantifikovatelné (je možné převést je na jednotnou škálu, která je pak na mapě vyjádřena škálou barevnou, barevnou sytostí apod.). Vedle toho však existuje řada komplexnějších jazykových jevů, které jsou závislé na prostoru, a tím si přímo říkají o mapové zpracování, ale v dvourozměrné mapě není možné zachytit je přehledně a jasně. Typicky jsou takovými jevy jazykové (gramatické) subsystemy, zahrnující více prvků a jejich vzájemné vztahy (např. systém krátkých vokálů). Běžným postupem je rozložit tyto komplexy na jednotlivé prvky a ty zmapovat odděleně. Uživatelé map, který by chtěl znát celý komplex (systém krátkých vokálů) v několika bodech mapy, pak nezbyvá než projít všechny tyto mapy ve všech těchto bodech a informace si z nich, s pomocí „tužky a papíru“, sestavit.

Brzy však něco takového bude záležitostí několika kliknutí myší, alespoň v rámci české dialektologie. Ve spolupráci dialektologického oddělení Ústavu pro jazyk český a Katedry geoinformatiky Univerzity Palackého v Olomouci totiž vzniká projekt nářečního geoportálu, kde by jednou z inovativních funkcí vyvíjeného interaktivního mapového prostředí měla být právě možnost snadného získávání takovýchto komplexních gramatických informací. Projekt se v první fázi zaměří na hláskosloví a na substantivní a adjektivní deklinaci, materiálově vyjde z výzkumů pro Český jazykový atlas (ČJA), ale obohatí tento materiál z řady dalších zdrojů, takže nová data budou v převaze. Na geoportálu vzniknou v této fázi řádově stovky map, týkajících se nářeční gramatiky, které budou vybaveny řadou nástrojů a funkcí. V mapovém prostředí např. bude možné srovnávat zvolené mapy a překládat je přes sebe. Nové zpracování dialektologických dat umožní automaticky modelovat průběhy izoglos, dále spojovat již proběhlé systematické výzkumy (založené na síti výzkumných bodů) s relativně neuspořádanými nářečními daty, náhodně rozmístěnými v čase a prostoru, a s jejich pomocí tyto původní výzkumy zpřesňovat. Bude možné rekonstruovat nářeční stav i v lokalitách, kde

<sup>1</sup> Příznačný je z tohoto hlediska např. svazek mapových ukázek jazykově zeměpisného sborníku *Language and Space. Language Mapping* (LSLM II). Ve snaze autorů prezentovat zde specifika různých proveniencí a špičkové výsledky minulého i současného vývoje, se do jinak pozoruhodného výběru dostala řada map sice originálních a vysoce odborných, ale nepraktických, s nimiž se zřejmě dá pracovat jen velmi pomalu a obtížně (z nejmarkantnějších: LSLM II, m. 607, 1102, 1306, 1308, 2607, 3102, 3103).

<sup>2</sup> *Eye-tracking*, sledování pohybů očí, poskytuje kartografům zpětnou vazbu o percepci mapy. Testování mapy probíhá tak, že pokusným osobám je zadáno několik úkolů, při nichž mají v mapě najít a vyčíst z ní různé typy informací, které má mapa sdělovat. Při tom jsou sledovány pohyby očí po testované mapě a je měřen čas. Cílem je (mnohdy po několika cyklech testování, vyhodnocování a úprav) dosáhnout mapového zobrazení, které je bezprostředně čitelné a intuitivně srozumitelné. Jedním z průkopnických pracovišť u nás i ve světě je na tomto poli Katedra geoinformatiky Univerzity Palackého v Olomouci (Popelka et al. 2012, 104–106, 112–113; Popelka, Brychtová, Voženilek 2012, 72–73, 79–83), s níž na vývoji nových dialektologických map úzce spolupracujeme.

žádný výzkum neproběhl, stanovovat kritické body výzkumných sítí a tím zefektivnit a lépe cílit terénní výzkum. Především však tyto nástroje umožní nahlédnout předevo nářečních gramatických vztahů z úhlů pohledu, které pro svou pracnost zůstaly dosud neuskutečny.

### Atomický jev a gramatický komplex

Vyděme však nejdřív z vymezení dvou pojmů, které nám v této studii usnadní vyjadřování: atomický jev a gramatický komplex. Jsou to pojmy, s nimiž jsme tu už neformálně (v parafrázích) pracovali. Jako gramatický komplex můžeme chápat gramatickou strukturu, která je provázána vzájemnými vztahy a je relativně samostatná (dobře definovatelná a oddělitelná) v rámci jazykového systému. Takovýmto gramatickým komplexem může být např. řečený systém krátkých vokálů, deklinační subparadigma imperativu slovesa *jít*, soubor všech koncovek nom. pl. životných maskulin apod. Tento gramatický komplex se pak skládá z atomických jevů, které definujeme jako relativně nesložené prvky gramatických komplexů či jazykového systému nebo též jako hodnotu nějaké jednoznačně definované proměnné: např. střídnice za pův. *i* po tupých sykavkách, gen. sg. substantiva *vejce* apod.

Takovéto atomické nářeční jevy se dají snadno zmapovat, protože jeden bod na mapě nám přiděluje ideálně jednu hodnotu, případně dubletu, tripletu apod. U gramatických komplexů je mapové zobrazení komplikované nejen proto, že každému bodu mapy přísluší větší množství proměnných, tudíž i větší množství hodnot, ale i proto, že mezi proměnnými či hodnotami jsou vzájemné vztahy, které je třeba graficky znázornit.

Cesta k novému mapovému zpracování gramatických komplexů přitom vede přes specifické uchopení atomických jevů, které navazuje na tradici ČJA.

### Pravidelné regionální obměny (PRO)

Český jazykový atlas měl tu paradoxní výhodu, že vznikl jako jeden z posledních velkých atlasů národního jazyka, ať už mezi jazyky slovanskými (srov. Kloferová 2010, 319–329; 2015, 44), nebo i v rámci evropských jazyků, které takovýto velký atlas mají (srov. Scheuringer 2010, 160–164; Kruijssen, van der Sijs 2010, 185–190; Edlund 2010, 203–204; Ramisch 2010, 240–242; Swiggers 2010, 279–291). Mohl tak využít předchozích zkušeností, poučit se z metod a výsledků jiných zpracování, posoudit jejich přednosti a nedostatky. Jako jeden z mála atlasů volí náročnější a lépe čitelné syntetické mapování proti mapám analytickým. Data z výzkumné sítě tudíž nepřenáší na mapu prostřednictvím bodových značek (co značka, to hodnota), ale díky abstraktnějšímu zpracování materiálu tak činí prostřednictvím ploch (skupina sousedících shodných hodnot vytváří plošně vyznačený areál)<sup>3</sup>. Takových jazykových

<sup>3</sup> Bodové značky jsou v ČJA též přítomny, ale pouze jako doplňkový zobrazovací prostředek (srov. Utěšený 1974; ČJA I, 36–38). Podrobněji rozdíl mezi mapou syntetickou a analytickou vymezuje J. Pravda (2002, 375–377).

atlasů existuje i v evropském a světovém kontextu jen nemnoho, jmenujme např. raný dánský jazykový atlas *Kort over de Danske Folkemål med Forklaringer* (KDF), atlas britské angličtiny *An Atlas of English Dialects* (AED), německý *Deutscher Wortatlas* (DW) a regionální *Nordbaierischer Sprachatlas* (NordbSA), mezi slovanskými jazyky pak atlas lužickosrbský *Sorbischer Sprachatlas* (SorbSA). I mezi nimi má však Český jazykový atlas jeden unikát, a to koncepci pravidelných regionálních obměn.

Tato koncepce využívá skutečnosti, že k některým hláskoslovným (a hláskoslovně morfologickým) změnám dochází pravidelně, téměř bez ohledu na podmínky a okolnosti v dalších jazykových plánech. Je tudíž možné od nich abstrahovat a vyčlenit je jako soubor samostatných map. Např. změna *ou* v *ó* má jednoznačnou izoglosu, která nám shodně prostorově rozděluje dvojice: *housenka* : *hósenka*, *zamknout* : *zamknót*, *dížou* : *dížó* aj. Logickým krokem bylo nemapovat tuto hranici pokaždé znovu při mapování těchto a dalších jevů, ale vyčlenit ji jako samostatnou mapu a z map všech ostatních jevů tuto hranici odebrat. Každá ze zmíněných dvojic tak po této abstrakci představuje jednu hodnotu a v mapě tvoří jednu nepředělenou plochu, jeden areál<sup>4</sup>. Tímto způsobem se složené jevy rozložily na jevy atomické, lépe mapovatelné. Když se chceme zpětně dobrat konkrétní realizace zobrazeného jevu v určitém bodě mapy, musíme do jednoho atomického jevu (*zamknout*) dosadit jiné atomické jevy jakožto proměnné (zde *ou* > *ó*; v ČJA jsou tyto proměnné značeny podtržením *zamknout*). Nejde-li nám o jeden bod, ale o celou mapu, pak do jedné mapy dosazujeme druhou, překládáme je přes sebe, ovšem pouze v areálech, jejichž hodnota danou proměnnou obsahuje (jestliže na mapě figurují dva areály, *zamknout* a *zamčít*, do každého z nich vkládáme jinou izoglosu na základě rozdílných proměnných). Způsob skládání či vkládání map a izoglos však nebyl v ČJA nijak prakticky vyřešen a zpětné spojování atomických jevů tak bylo ergonomicky náročné.

V ČJA byly pravidelné regionální obměny pojaty jakožto binární opozice (izoglosa rozděluje mapu vždy na dva areály, v jednom z nich je jev přítomen, ve druhém není). V teoretickém výkladu PRO je uvedeno, že izoglosy „na Moravě představují víceméně přesné hranice hláskoslovných obměn, v Čechách jde vesměs o izoglosy ohraničující oblasti maximálního potenciálního výskytu sledovaného jevu“ (ČJA I, 49). V konfrontaci s nekompromisností a důsledností počítačových algoritmů se však ukazuje, že nejvhodnější je rozdělit PRO, nehledě k jejich lokalitě, do tří kategorií podle způsobu platnosti mapovaného jevu na: jev důsledný (konsekvence), nedůsledný (fluktuace) a ustupující (degrese). Jev důsledný, konsekventní lze považovat v celém svém územním rozsahu za systémový, tedy pravidelný a očekávatelný. Jev nedůsledný se ve svém územním rozsahu objevuje příležitostně a nepravidelně, tedy může se se značnou pravděpodobností neobjevit, i když jsou splněny všechny hláskoslovné podmínky, za nichž se obvykle objevuje (fluktuace tak tvoří neostrý přechod mezi pravidelnou a nepravidelnou nářeční obměnou; nepravidelná je už lexikalizovaná, nedůsledná pravidelná má v jazykovém systému i při své konkrétní absenci jistou latenci, možnost se objevit). Jev ustupující, regresivní se pak objevuje v části území důsledně a v části nedůsledně, zpravidla má své jádro vzdálené od izoglosy a vytrácí se směrem k ní.

<sup>4</sup> Z mapy tvaru instr. sg. slova díže fakticky odebrána nebyla, v morfologických mapách byly nakonec hláskoslovné varianty často zakreslovány čárkovaně (ČJA IV, 151), ale zde nám jde o principy zobrazení.

Vrátíme-li se k pojetí PRO jako logické proměnné<sup>5</sup>, je pak možné kvantifikovat, s jakou pravděpodobností nabývá proměnná hodnoty příznakového členu binární opozice v každém bodě výzkumné sítě. Jinými slovy: je možné na mapě zobrazit míru pravděpodobnosti té které PRO. Máme-li výzkumem pro ČJA zhruba podchycenou základní slovní zásobu na celém tradičně osídleném území, je možné na tomto vzorku vypočítat a zmapovat, s jakou pravděpodobností či důsledností se kde která pravidelná regionální obměna vyskytuje. V případě fluktuace a degrese pak není nutné vykreslovat izoglosu jen v oblasti maximálního rozsahu výskytu jevu, ale je pro ni možné stanovit jiné kritérium, případně zobrazit zóny různé míry pravděpodobnosti zhruba podobným způsobem jako terénní vrstevnice. Takovéto dialektometrické možnosti už nebyly v rámci ČJA zpracovány, nabízí se však ke zpracování nyní.

### Interpolace

Díky pravidelným regionálním obměnám je v ČJA dosaženo toho, že jazykové jevy jsou zpravidla rozložené na jevy opravdu atomické. Zmapované izoglosy tak přísluší vždy právě jednomu jevu, abstrahovanému od jevů dalších. Takovýto přístup se nyní, v kontextu počítačového zpracování, ukazuje jako mimořádně šťastný. A to především s ohledem na nové interpolační vymezení průběhu izoglos a vůbec interpolační zpracování veškerých dat.

Interpolace je v kartografii metoda prostorového modelování, kterou se v mapách dopočítávají data na místech, kde tato data nemohla být zjištěna exaktně. V řadě případů, kdy se získává hodnota nějaké proměnné z většího území, totiž není možné získat tuto hodnotu z každého myslitelného bodu na mapě. Vychází se z omezeného množství vyzkoumaných (změřených, sondovaných) bodů a zbylé oblasti mezi nimi se domodelují na základě prostorových vztahů mezi vyzkoumanými body, zjištěných hodnot v těchto bodech i na základě dalších faktorů.

Typickým příkladem použití interpolace je mapování nadmořské výšky, kterou není možné změřit ve všech potřebných bodech a průběhy vrstevnic a nově i 3D modely reliéfu a povrchových objektů se proto dopočítávají právě interpolačními metodami. Dalším zajímavým využitím, spojeným s jednou specifickou interpolační technikou zvanou kriging, je výpočet tvaru a průběhu zlaté žíly na základě sond (důlních vrtů), jejich prostorového rozložení a množství zlata v nich. Využití interpolací v geologických mapách (půdní mapy, mapy nerostných ložisek, mapy půdního sucha apod.) je zcela typické, neboť omezený počet sond je zde nevyhnutelný. Aplikaci interpolací nacházíme i v meteorologických mapách tam, kde vycházejí ze sítě meteorologických stanic (mapy srážek, teplotní mapy, větrné mapy) a obecně v řadě případů syntetického tematického mapování.

Je zjevné, že jako dialektologové jsme se svou sítí výzkumných bodů ve velmi podobné situaci. Rovněž máme omezené množství bodových dat a kompletní zmapování všech myslitelných bodů (obcí a částí obcí) je dalece mimo možnosti jakéhokoli

---

<sup>5</sup> Tedy proměnné, která nabývá pouze dvou hodnot (ano/ne), podle toho, zda se pravidelná regionální obměna v konkrétním případě realizuje, nebo ne.

reálného výzkumného týmu. Interpolace se tak nabízí jako postup uplatnitelný i v dialektologii. Také ostatně nejsme první, koho napadlo tuto metodu v dialektologických mapách využít (srov. např. Sibler et al. 2012; Scherrer 2012, 123–137; Grieve 2013a; Grieve 2013b; Schwarz 2014; Thebpanya, Hatfield 2016, Chagnaud et al. 2017), převážně se však užití interpolací omezuje na dialektometrickou práci s kvantitativními a agregovanými daty, což je nyní v tomto dynamicky se rozvíjejícím se oboru trendem. Pouze první a poslední studie z výše uvedených pracují s jednotlivými rozlišenými jazykovými jevy a pouze poslední tak činí na základě kvalitativních dat, nikoli kvantitativních (procentuálního zastoupení ve vzorku), což vyžaduje zapojení méně tradičních interpolačních technik. Interdisciplinární spolupráce dialektologického oddělení Ústavu pro jazyk český s Katedrou geoinformatiky Univerzity Palackého nám dává v aplikaci interpolačních technik velmi široké možnosti.

Obě pracoviště ve spolupráci připravují interpolační schéma, které u atomických jevů zohlední nejen vlastní prostorové uspořádání vstupních dat, ale i další modifikující faktory prostředí jako jsou přírodní překážky nebo historické správní a politické hranice (hranice farností a panství, jejichž váha bude stanovena na základě jejich agregované diachronní stability). Interpolační schéma je možné ladit nastavováním vah a koeficientů jednotlivých zúčastněných faktorů a toto ladění bude rozsáhle experimentálně ověřováno na trénovacích datech pro dosažení co nejvyšší věrnosti modelování.

Vzniknou tak modelované průběhy izoglos na přesném a přibližitelném zeměpisném topopodkladu, včetně vyjádření jejich nejistoty prostřednictvím vypočteného rozostření linie či barevného přechodu (blíže k metodám vizualizace nejistoty viz např. Kinkeldey, MacEachern, Schiewe 2014, 378–384; Brus 2014, 69–74). Tento problém nemusely tradiční dialektologické mapy řešit, neboť u nich topopodklad vesměs chybí; jedná se o „slepé“ mapy s vyznačenými vybranými body bez možnosti přiblížení nebo identifikace přesného místa. Tudíž linie izoglosy může být vedena kdekoli mezi dvěma body aniž by to jakýmkoli způsobem měnilo informace extrahovatelné z mapy. To platí i o mapách v ČJA. Prostřednictvím interpolací však bude možné získat ze stejných dat více informací a lepší mapové zobrazení, s podkladem mapy zeměpisné i různých map historických.

Vytvoření takovýchto map (jakéhosi zdokonaleného zobrazení již jednou zmapovaného materiálu) ovšem není konečným cílem projektu, ale pouze mezikrokem. Při našem postupu se totiž interpolace nebude odehrávat pouze na mapách a pouze v prostoru kolem izoglos, ale na veškerých prostorových datech. Každý atomický jev bude takto modelován pro všechny body tradičního teritoriálního dialektu (tj. části obcí). Zajišťovat to bude vyvíjený software ProMAP, který bude vybaven funkcemi, jež umožní nejen data interpolovat, ale už předtím formovat a předzpracovat: poloautomaticky analyzovat a upravovat vstupní data na atomické jevy a přidávat každé jednotlivé hodnotě atomického jevu gramatické a jiné atributy, které později umožní systematicky uspořádávat atomické jevy do gramatických komplexů. Jakákoli průběžně dodávaná data posílí a zpřesní data již zpracovaná, model se bude pravidelně znovu přepočítávat se zohledněním nových údajů a mapy a průběhy izoglos se tak budou dále zpřesňovat.

## Software ProMAP

Když vznikaly jazykové atlasy, bylo zpravidla dbáno na to, aby výzkumná síť byla relativně rovnoměrně rozložena, aby výzkum proběhl v co nejkratší době (byl relativně synchronní), aby data sbíraná na různých místech byla jednotná a vzájemně porovnatelná (jednotný dotazník), výzkum pro ČJA byl navíc dosti ojedinelý v tom, že všechna data zaznamenávali pouze dialektologové-odborníci (srov. Voráč, Racková 1969, 313–316; Kloferová 2007, 359–360). Všechny tyto faktory byly v dané době nesmírně důležité a zkvalitňovaly výsledek. Zpracujeme-li ovšem získaná data prostřednictvím interpolací, mnohé z těchto požadavků na nasbíraná data už nejsou nutné, aniž by to přitom snižovalo kvalitu výsledku. Právě naopak.

Interpolace umožňují pracovat se zcela náhodně prostorově rozmístěnými bodovými daty (tzv. *scattered data*; srov. Ježek 2015, 13). Jejich rovnoměrné rozložení je samozřejmě stále výhodné, ale není nutné, modelování ho nepotřebuje. Bylo tak například možné zpracovat dva nezávislé dialektometrické výzkumy se zcela odlišně rozloženou výzkumnou sítí a modelovat z nich dvě velmi podobné mapy (Grieve 2013a). Ani časový požadavek synchronního výzkumu už není tak striktní, neboť časová vzdálenost (od nějakého zvoleného časového bodu) může přidělovat zjištěné hodnotě váhu, s níž se bude promítat do výsledků interpolace (při dostatečném množství dat je možné mapu i rozpohybovat, přidat jí časový rozměr, přičemž rozdílný výsledek interpolace v různých časových bodech bude záležitostí pouze změn časových vah u jinak stále stejné sady vstupních dat). Nezbytný vlastně není ani jednotný dotazník, protože rozvolněním požadavků rovnoměrné bodové sítě a synchronicity, je možné data k jednotlivým zkoumaným jevům sbírat z jakéhokoli relevantního zdroje (z nějž jsme s to zjistit hodnotu atomického jevu, lokalitu a časové určení a potvrdit autochtonnost). Dotazníkové šetření prováděné školeným dialektologem tím sice není zcela zastupitelné (zvláště pak u jevů běžně nezaznamenávaných nebo obtížně rozlišovaných, které není možné zmapovat bez systematického zacílení), na druhou stranu v dostatečně rozsáhlém prohledávatelném korpusu nářečních textů (jako je Archiv souvislých nářečních textů na dialektologickém oddělení), lze většinu myslitelných jevů najít v dostatečném množství, aniž se na ně při sběru kdokoli zaměřoval. Původní výzkum ČJA vycházející ze sítě necelé pětistovky bodů tak bude obohacen body dalšími, u každého jevu pochopitelně jiným počtem a v jiném prostorovém rozložení. Pro každý jev tedy bude síť referenčních bodů jiná, přičemž budou přidány i další jevy, které původní výzkum ČJA nezaznamenával a které tudíž nebudou opřené o jeho výzkumnou síť.

V první fázi se budou mapovat jevy hláskoslovné a morfologické (substantivní a adjektivní deklinace). Ty budou pocházet zhruba ze čtyř skupin zdrojů: z původního dotazníkového šetření pro ČJA, z bohatého korpusu nářečních monografií, studií a studentských prací zaznamenávajících hláskosloví a morfologii v konkrétních lokalitách, dále ze zpracovaných terénních výzkumů dialektologického oddělení a nakonec z rozsáhlého korpusu nářečních textů.

Z těchto dat bude vznikat prostorová databáze prostřednictvím softwaru ProMAP. Jak už bylo řečeno, právě v tomto softwaru budou probíhat interpolace. Na základě vstupních dat ke každému jednomu atomickému jevu budou domodelovány hodnoty ve všech dalších prostorových bodech, kde vstupy budou chybět. Software však bude také

schopen poloautomaticky převádět zadávaná surová data na atomické jevy. Tedy po zadání sady neupravených hodnot jednoho složeného jevu (např. různé tvary a nářeční formy slova *býk*: *bejci*, *békama*, *bikom*, *býkamy*, *bykoji*..., u nichž bude určen mj. tvar, lokalita a lemma *býk*) bude software schopen vytipovat pravidelné regionální obměny, morfologický šev mezi koncovkou a zbytkem slova a morfonémovou alternaci, vytvořit abstraktní podobu tvaru i samotné koncovky. Bude se tak dít algoritmizovanou sérií komparací napříč daty, ale výsledky budou procházet ještě lidskou kontrolou. Veškerá data takto budou rozkládána na atomické jevy a podrobována interpolacím.

Každá entita (konkrétní hláska, konkrétní tvar) bude v databázi označena maximem gramatických atributů, ať už půjde o místo a způsob artikulace a zvukové kvality u fonémů nebo gramatické kategorie a přítomnost morfonémových alternací u tvarů. Mimoto budou sledovány i další atributy jako původ hlásky, lemma tvaru, deklinační typ, dubletnost, pádová homonymie apod. Na základě těchto atributů budou moci být později jednotlivé atomické jevy usouvztažňovány.

Vlastní zpětný sklad atomických jevů do gramatických komplexů se převážně nebude odehrávat v softwaru ProMAP (byť některá spojení atomických jevů budou v zájmu hladkého chodu všech nástrojů předpřipravena už v jím vytvořené databázi), ale povětšinou se budou komponovat teprve v prezentační části celého projektu, v geoportálu DiaMa.

### Geoportál DiaMa

Geoportál DiaMa bude online sbírkou velkého množství interaktivních nářečních map (v první fázi jich budou řádově stovky) s možností vyhledávání, přiblížení, s volitelnými topopodklady a s pokročilou funkcionalitou. Mapy budou primárně zobrazovat jednotlivé atomické jevy, grafické pojetí zobrazovaných areálů bude důsledně plošné. Součástí map bude i multimediální exemplifikační obsah v zobrazitelných bodových značkách, o němž však v této studii nehovoříme, neboť nesouvisí přímo s naším tématem.

Vzhledem k důsledným interpolacím budou v každé mapě všechny body tradičního osídlení nabývat nějakou zjizitelnou hodnotu (u interpolovaných bodů to bude nejen hodnota kvalitativní, ale i vyjádření pravděpodobnosti této kvalitativní hodnoty). Každý prostorový bod na vymezeném území tedy bude vykazovat hodnotu pro všechny atomické jevy, napříč všemi mapami.

Součástí interaktivního prostředí geoportálu bude datový prohlížeč, který umožní uživateli přehledně zobrazit data, z nichž jsou mapy generovány. To je věc, která zatím v dosavadních dialektometrických a jiných pokročileji počítačově mapovaných výsledcích velmi chybí. Datový prohlížeč bude fungovat ve formě postranního panelu vedle mapy a např. po vyhledání konkrétní obce či části obce bude schopen zobrazit veškeré nářeční údaje k tomuto bodu, včetně interpolovaných (ty budou opatřeny pravděpodobností v indexu u hodnoty). Především však nabídne uživateli možnost vytvořit si z dostupných dat libovolnou datovou sestavu. Některé sestavy (typu deklinační paradigma) budou předpřipravené, jiné, uživatelsky volitelné, budou mít k dispozici několik prostorových vzorců či šablon, jejichž prostřednictvím bude možné data uspořádat



a vizualizovat. Posune-li pak uživatel ukazatelem v mapě<sup>6</sup>, konfigurace datové sestavy zůstane, ale změní se v ní hodnoty proměnných, případně i vazby mezi nimi. Na topopodkladu bude přitom možné zobrazit přes sebe přeložené mapy všech atomických jevů obsažených v datové sestavě, nicméně ve zjednodušeném zobrazení (linie izoglos namísto barevných ploch areálů). Přetažením posuvného ukazatele přes některou z linií bude fungovat jako její odečtení, neboť jedna z proměnných v postranním panelu při tom změní hodnotu. Vznikne tím jakási interaktivní legenda této svodné mapy a současně náhrada mapových popisků, které by v tak složité (a nadto volitelné) mapě nefungovaly. Ke zpřehlednění bude zřejmě zapotřebí, aby se každá proměnná v zobrazeném gramatickém komplexu barevně shodovala s odpovídajícími liniemi izoglos (tj. izoglosy pocházející z mapy jednoho atomického jevu budou mít všechny stejnou barvu, totožnou i s barvou písma proměnné reprezentující tento atomický jev v gramatickém komplexu). Další zpřehledňující grafické funkce (zvýrazňování, odebrání izoglos apod.) budou teprve předmětem dalšího vývoje.

Překládat libovolné mapy přes sebe bude možné i bez vytvoření specifické datové sestavy, stejně tak porovnávat mapy vedle sebe nebo více map rozložených dlaždicově. Mapové prostředí bude mít řadu dalších funkcí včetně možnosti zapnout a vypnout sít' výzkumných bodů, odlišit v ní body původního systematického výzkumu od scattered data atd. Mapu naprosté většiny jevů bude také možné zobrazovat na různých úrovních abstrakce – jakožto atomický jev i v kombinaci s proměnnými, od nichž je u atomických jevů odhlíženo.

Velmi důležitou funkcí pak bude možnost u každého bodu mapy vygenerovat všechny jevy pro které platí, že se daný bod nachází v blízkosti nějaké izoglosy. Tuto funkci je možné využít např. při plánování terénního výzkumu v daném místě (vybírání jevy, na něž je v tomto bodě vhodné se v rámci ověřovacího výzkumu zaměřit). Na základě interpolací je také možné vytipovat "hluchá místa" bodových sítí, a to např. tam, kde se stýkají více než dva areály nebo kde hustota sítě je příliš nízká na proměnlivost hodnot a říká si o zpřesnění. Geoportál tedy pravděpodobně umožní i lépe zacílit a zefektivnit terénní výzkumy, zaměřit se na správných místech na ta správná data a vynechat sběr dat skýtající podstatně nižší pravděpodobnost údaje, který by zpřesňoval či obohacoval naše znalosti.

## Literatura

- AED, Upton C., Widdowson, J. D. A., *An Atlas of English Dialects*, Oxford 1996.
- Brus J. (2014), *Vizualizace nejistoty v environmentálních studiích*, Olomouc.
- ČJA, Balhar J. et al., *Český jazykový atlas*, sv. 1–5, Dodatky, Praha, 1992–2011. <https://cja.ujc.cas.cz/>, 10.08.2020.
- DW, von Mitzka W., Schmitt L. E., *Deutscher Wortatlas*, Bd. 1–22, Gießen 1951–1980.
- Eldlund L.-E. (2010), *Mapping the North Germanic Languages*, [in:] *Language and Space. Language Mapping*, vol. I–II, eds A. Lameli, R. Kehrein, S. Rabanus, Berlin – New York, s. 203–237.

<sup>6</sup> Tento ukazatel např. po vyhledání označí vyhledané místo, nejde o kurzor myši.

- Grieve J. (2013a), *A statistical comparison of regional phonetic and lexical variation in American English*, „Literary and Linguistic Computing“, no. 1, s. 82–107.
- Grieve J. (2013b), *Ordinary kriging in dialectology*, [in:] *Proceedings of the 5th Conference on Quantitative Investigations in Theoretical Linguistics. QITL-5*, eds T. Wielfaert, K. Heylen, D. Speelman, Leuven, s. 23–24.
- Chagnaud C. et al. (2017), *ShinyDialect. A cartographic tool for spatial interpolation of geolinguistic data*, [in:] *GeoHumanities'17. Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*, New York, s. 23–30.
- Ježek J. (2015), *Geostatistika a prostorová interpolace*, Praha.
- KDF, Bennike V., Kristensen M., *Kort over de Danske Folkemål med Forklaringer*, vol. 1–2, København 1898–1912.
- Kinkeldey C., MacEachern A. M., Schiewe J. (2014), *How to assess visual communication of uncertainty? A systematic review of geospatial uncertainty visualisation user studies*, „The Cartographic Journal“, no. 4, s. 372–386.
- Kloferová S. (2007), *Dialektologie*, [in:] *Kapitoly z dějin české jazykové a bohemistiky*, ed. J. Pleskalová et al., Praha, s. 336–376.
- Kloferová S. (2010), *Mapping the Slavic languages*, [in:] *Language and Space. Language Mapping*, vol. I–II, eds A. Lameli, R. Kehrein, S. Rabanus, Berlin – New York, s. 317–332.
- Kloferová S. (2015), *Národní atlasy slovanských jazyků: historie a současnost*, „Gwary Dziś“, t. 7, s. 43–55.
- Kruijssen J., van der Sijs N. (2010), *Mapping Dutch and Flemish*, [in:] *Language and Space. Language Mapping*, vol. I–II, eds A. Lameli, R. Kehrein, S. Rabanus, Berlin – New York, s. 180–202.
- LSLM, *Language and Space. Language Mapping*, vol. I–II, eds A. Lameli, R. Kehrein, S. Rabanus, Berlin – New York 2010.
- NordbSA, Gütter A., *Nordbairischer Sprachatlas*, München 1971.
- Popelka S. et al. (2012), *Advanced Map Optimization Based on Eye-Tracking*, [in:] *Cartography – A Tool for Spatial Analysis*, ed. C. Bateira, Rijeka, s. 99–118.
- Popelka S., Brychtová A., Voženilek V. (2012), *Eye-tracking a jeho využití při hodnocení map*, „Geografický časopis“, č.1, s. 71–87.
- Pravda J. (2002), *Analytické a syntetické mapy v klasifikácii máp*, „Geografický časopis“, č. 4, s. 367–380.
- Ramisch H. (2010), *Mapping British English*, [in:] *Language and Space. Language Mapping*, vol. I–II, eds A. Lameli, R. Kehrein, S. Rabanus, Berlin – New York, s. 238–252.
- Scherrer Y. (2012), *Generating Swiss German sentences from Standard German: a multi-dialectal approach*, Genève.
- Scheuringer H. (2010), *Mapping the German language*, [in:] *Language and Space. Language Mapping*, vol. I–II, eds A. Lameli, R. Kehrein, S. Rabanus, Berlin – New York, s. 158–179.
- Schwarz C. (2014), *Conservative and innovative dialect areas*, „Taal en Tongval“, no. 1, s. 65–83.
- Sibler P. et al. (2012), *Cartographic Visualization in Support of Dialectology*, [in:] *Proceedings – AutoCarto 2012*, Columbus, [https://www.cartogis.org/docs/proceedings/2012/Sibler\\_etal\\_AutoCarto2012.pdf](https://www.cartogis.org/docs/proceedings/2012/Sibler_etal_AutoCarto2012.pdf), 15.08.2020.
- SorbSA, Faßke H., Jentsch H., Michalk, S., *Sorbischer Sprachatlas*, Bd. 1–15, Bautzen 1965–1996.

- Swiggers P. (2010), *Mapping the Romance languages of Europe*, [in:] *Language and Space. Language Mapping*, vol. I–II, eds A. Lameli, R. Kehrein, S. Rabanus, Berlin – New York, s. 269–300.
- Thebpanya P., Hatfield S. L. (2016), *Visualizing dialect variation on a 3-D interpolated map: A case study in Chiang Mai, Thailand*, „Papers in Applied Geography“, no. 2, s. 243–252.
- Utěšený S. (1974), *K problematice symbolů v lingvistickém kartografování*, „Slovo a slovesnost“, č. 4, s. 327–335.
- Voráč J., Racková I. (1969), *Práce na Českém jazykovém atlase*, „Slovo a slovesnost“, č. 3, s. 312–318.