

---

## ROZHLED

---

# Motivické a tematické klastry v básnických textech české poezie 19. a počátku 20. století

K NOVÝM MOŽNOSTEM VYUŽITÍ DATABÁZE  
ČESKÁ ELEKTRONICKÁ KNIHOVNA

Lucie Kořínková  — Tereza Nováková — Michal Kosák  —  
Jiří Flaišman  — Karel Klouda 

---

### **Motif and thematic clusters in 19th and early 20th century Czech poetic texts New ways of using the Czech Electronic Library database**

The aim of this article is to analyse the possibilities and above all the results of the machine processing of motif and thematic clustering in 19th century poetry, which was performed on a corpus of poems in the Czech Electronic Library full-text database.

---

### **Klíčová slova / Keywords**

Česká elektronická knihovna — plnotextová databáze — motivické klastry —  
tematické klastry — Top2Vec — poezie 19. století

Czech Electronic Library — full-text database — motif clusters — thematic  
clusters — Top2Vec — 19th century poetry

---

### **Kontakt**

Ústav pro českou literaturu AV ČR; korinkova@ucl.cas.cz — novakova@ucl.cas.cz —  
kosak@ucl.cas.cz — flaisman@ucl.cas.cz — Fakulta informačních technologií ČVUT;  
karel.klouda@fit.cvut.cz

Lucie Kořínková — Tereza Nováková — Michal Kosák — Jiří Flaišman — Karel Klouda.  
„Motivické a tematické klastry v básnických textech české poezie 19. a počátku 20. sto-  
letí. K novým možnostem využití databáze Česká elektronická knihovna“; *Česká lite-  
ratura* LXXII, č. 2, s. 204–217

<https://doi.org/10.51305/cl.2024.02.04>

Od doby vzniku České elektronické knihovny<sup>1</sup> (dále též ČEK) výrazně pokročila digitalizace českých knihovních fondů a mnohé z básnických knih, které byly dříve čtenářské obci přístupné jen obtížně, jsou dnes v digitálních skenech originálů k dispozici v repozitářích typu Kramerius. Přesto má databáze ČEK i v takto proměněných podmínkách své nezastupitelné místo: pro čtenářský zájem o literaturu 19. století plní i nadále funkci unikátní knihovny soustřeďující daný materiál na jednom místě; pro zájmy badatelské je zase zásadní, že jde o materiál ucelený, jednotně edičně ošetřený a relativně homogenní, neobsahující mimo českou poezii 19. a počátku 20. století již nic dalšího (s drobnou výjimkou odlišně databázově zpracovávaných doprovodných textů knih a případných nebásnických textů v almanaších).<sup>2</sup> Nad tímto materiálem navíc od počátku fungují digitální nástroje k jeho dalšímu třídění, slovníky, jednoduché statistiky a vyhledávání (včetně badatelsky velmi užitečného strukturovaného vyhledávání kombinace slov v předdefinovaném prostoru verše, strofy či básně).

Velmi rychlý vývoj digitálního prostředí v uplynulých dvou dekadách nutně způsobil zastarávání technických řešení databáze: software nebylo možno jednoduše upgradovat směrem k novým možnostem, jež se v oblasti digitálního zpracování textových dat v poslední dekádě výrazně rozšířily (mezi ty patří příkladně lemmatizované vyhledávání lexikálních jednotek), a v konečném důsledku vedlo zaostávání za vývojem i ke komplikacím s přístupností databáze pro širokou obec možných uživatelů.<sup>3</sup>

Cenná data shromážděná v ČEK je dnes tedy třeba nově databázově ošetřit, ale také k nim vytvořit nové uživatelské rozhraní, které stávající funkcionalitu

- 1 Vytvoření „Plnotextové databáze české poezie 19. a počátku 20. století“, jak níž podtitul České elektronické knihovny, bylo iniciováno v polovině devadesátých let 20. století (srov. SVADBOVÁ 2000). Projekt si kladl za cíl „shromáždit a zpřístupnit ve formě elektronické knihovny první vydání básnických textů od počátku novodobé česky psané poezie až po její plný rozmach, a tak uchovat a v elektronické podobě konzervovat celek české básnické produkce 19. století (v další fázi projektu pak poezii počátku století 20.)“. Rozsah plánované databáze byl z realizačních důvodů zúžen na knižní vydání básnických skladeb a sbírek (O PROJEKTU 2005). Výsledkem je databáze shromažďující vedle vybraných kolektivních almanachů zejména první vydání sbírek, respektive pozdější vydání týchž celků v básnických spisech či zásadních autorských přepracováních. Internetová podoba aplikace ČEK byla veřejnosti představena ve svém stávajícím datovém rozsahu čítajícím 1 700 sbírek roku 2005.
- 2 Dokladem badatelské užitečnosti dat shromážděných v ČEK je mezi jiným i to, že se tento materiál stal datovým základem pro vznik samostatné databáze Korpus českého verše a pro další projekty versologického týmu ÚČL pod vedením Petra Plecháče (pro více informací viz [www.versologie.cz](http://www.versologie.cz)).
- 3 Databáze ve své původní podobě sice stále ještě fungovala pro dříve vytvořené uživatelské účty, noví zájemci bez účtu se ovšem k jejím datům nedostali.

nejen plně zachovává, ale i rozšíří. Zároveň se nová aplikace (od července 2024 na známé webové adrese [www.ceska-poezie.cz](http://www.ceska-poezie.cz)) stane pokud možno mnohem přístupnější pro běžné čtenářské využití a konkrétní texty v ní budou vyhledatelné i přes dotazy v internetových prohlížečích. Nástroje digitálního zpracování dat i plánované didaktické extenze vybaví databázi tak, aby se z ní mohlo stát užívaný a užitečný nástroj badatelských aktivit, stejně jako výuky, a to nejen na bohemistikách vysokých škol, ale i na školách středních.

Vedle těchto dílčích cílů bylo základním úkolem rozšíření nástrojů databáze tak, aby se jednotlivé básně daly automaticky anotovat co do motiviky, v ideálním případě i co do nejvyšší sémantické roviny textu, tj. řečeno jazykem literární vědy naučit stroj rozeznávat motivy a téma básně.<sup>4</sup> Vstupními daty pro tzv. supervizované metody zpracování materiálu (tedy ty, kdy se algoritmus nejprve „učí“ na datech anotovaných lidmi) byly sady ručně označovaných básní; množina básní k anotování byla náhodně vygenerována z celku dat ČEK. U těchto textů tým literárních vědkyň a vědců z množiny tematických okruhů, jež si předtím stanovil, vybíral ke každé básni patřičnou tematickou kategorii, respektive křížení těchto kategorií.<sup>5</sup> Praktická činnost značkování dat pro potřeby strojového učení se úzce dotýkala problému svou povahou velmi literárněvědného, a sice jak definovat téma, jak motiv a jak přemýšlet o jejich vzájemném vztahu; nutnost přiřazování tematických labelů nás vedla k reflexi svých vlastních postupů při interpretaci významů básnického textu.<sup>6</sup>

Pro účely tohoto textu není nutné zpřítomňovat po léta trvající teoretickou diskuzi nad definicí motivu a tématu a jejich do značné míry komplikovaného vztahu,<sup>7</sup> pro řešitelky a řešitele projektu bylo východiskem uchopení

- 
- 4 Vlastní experimenty se strojovou analýzou se odehrávaly v rámci projektu TAČR odehrávaly na Katedře aplikované matematiky ČVUT pod vedením Karla Kloudy.
  - 5 Tematických okruhů jsme ve vzájemném konsenzu stanovili celkem 25: u části z nich jsme si byli se znalostí materiálu jisti, že budou (často hojně) zastoupeny, další byly stanoveny mimo jiné s ohledem na zásadní výklady (srov. např. MACURA 1999) kulturních a společenských fenoménů dlouhého 19. století z posledních desetiletí. Toto je jejich konkrétní výčet: dětství/mateřství/rodičovství; domov; exotika/cestovatelství; chudoba/útisk; náboženství/víra; pokrok/technika; politika/společnost; práce; příroda; revoluce/osvobození; samota/odcizení; smrt/umírání/stáří; tělo/sport; tradice/folklor; umění/literatura, poezie/tvorba; válka; vzdělání/škola; zábava/volný čas; zločin/trest; láska/milostný vztah; město; venkov; národ/vlast; historie; sebereflexe/vnitřní život.
  - 6 Šestičlenný tým ÚČL tímto způsobem popsal takřka dva tisíce básnických textů, nechtěným výsledkem bylo mj. též zjištění o nekonzistenci hodnocení v čase i ve srovnání mezi jednotlivými hodnotitelkami a hodnotiteli.
  - 7 K pojmovým definicím srov. též např. příslušná hesla *motiv* a *téma* ve *Slovníku literárněvědného strukturalismu* (SLÁDEK — FOŘT 2018: 472–477 a FOŘT 2018: 751–754), či ve *Slovníku novější literární teorie* (ŠIDÁK 2012: 327–330).

textu jakožto hierarchické struktury, v níž motivy jsou tradičně chápány jako nejmenší sémantické jednotky, diferencované dále např. významově či funkčně, a téma jakožto „abstrahovaná základní myšlenka textu“ (LUBKOLL 2006: 530) vytvářená motivickými trsy.

Častým jevem je situace, kdy je téma básně lexikálně průběžně signalizováno jen minimálně, většina textu je vystavěna z motiviky z jiné sémantické oblasti a hlavní téma se vyjeví až v samotném závěru básně. Stanovit v takovém případě hlavní téma čtenáři s interpretační zkušeností problém nečiní, je ovšem otázka, nakolik je v takovém případě toto hlavní téma sémanticky posouváno onou užitou motivikou. Stručně řečeno, nelze říci, že by básně o smrti s motivikou podzimního přírodního dění nebyla rovněž alespoň částečně básní přírodní lyriky.

Interpretačně ještě daleko komplikovanější situace může nastat v případech, kdy je téma motivice hodně vzdálené a zároveň na jeho lexikální manifestaci přímo ve verších dané básně vůbec nedojde — motivika je pak jen šifrou, ke které je třeba nalézt klíč v oblasti širšího kontextu: tento klíč může poskytnout titul básně, její dedikace či motto; do interpretační významových rovin básně při snaze o porozumění básnickému textu navíc vstupují také jemnější přediva vazeb intertextových,<sup>8</sup> informace o umístění básně v rámci sbírky (se zvláštním zvýznamněním počáteční a koncové pozice v rámci celku), ale také literárněhistorické informace o autorovi textu a o jeho zapojení do literárního provozu.

Experimenty se supervizovanými metodami strojového zpracování básní ČEK lze shrnout do závěru, že stanovení tématu v literárněvědném smyslu u uměleckého textu představuje pro strojovou analýzu (alespoň zatím) úkol příliš složitý, zjevně proto, že jde v každém jednom případě o úlohu interpretační s vyhodnocováním příliš mnoha proměnných, jejichž důležitost je třeba vážit podle situace. Případy, kdy klíč k tématu básně leží mimo vlastní text jejích veršů, pak použitými automatickými metodami úspěšně vyřešit v zásadě již z definice nebylo možné, neboť texty vstupující do těchto strojových analýz byly vždy zbaveny svých metadat (autorství, název, rok vydání či zařazení do celku sbírky tak pro algoritmus nehrají žádnou roli). Ještě před vlastní analýzou projde každý text krom očistění od metadat také takzvanou tokenizací dat: z textu je odstraněna interpunkce, velká písmena na počátku vět, ale také všechna neplnovýznamová slova („STOP words“); zbylá plnovýznamová slova jsou posléze lemmatizována a do strojové analýzy vstupují jako jednotlivé „tokeny“ v podobě svého základního tvaru. Všechny použité metody využívají techniky tzv. vnoření

<sup>8</sup> Tato intertextová navazování se mohou dít i zcela subtilně příkladně na rovině metrické, jak na ilustrativních příkladech českých básnických textů z přelomu 19. a 20. století dokládá nedávná monografie Roberta Kolára (KOLÁR 2023).

(angl. *embedding*), resp. vektorizace, kdy zkoumaný text převádějí na číselné vektory.

První sadu experimentů s automatickým rozeznáváním tématu nad daty původně z databáze ČEK provedl v rámci své diplomové práce na Katedře aplikované matematiky Fakulty informačních technologií ČVUT Martin Bendík (BENDÍK 2023). Cílem jeho práce bylo nad tímto materiálem vyzkoušet a následně zhodnotit jednotlivé metody užívané k automatické obsahové analýze textů. Při reflexi rizik, s nimiž se může strojová analýza básnického materiálu potýkat, se autor již na počátku svého úkolu dotkl mezi jiným problému poetického jazyka, zejména jeho figurativnosti a zvláštností lexika i gramatiky a slovosledu, jež se výrazně odchyľují od jazyka textů, na nichž jsou existující algoritmy strojové analýzy obsahu obvykle trénovány. Problém byl dopředu shledáván potenciálně také v možné nekoherenci dat<sup>9</sup> (IBID.: 5). Naopak zcela stranou pozornosti byla ponechána skutečnost, že téma coby pomyslný svorník významu básnického textu zkrátka nelze chápat jako prostou sumu nižších významových jednotek textu.<sup>10</sup>

Ze zmíněného základního diskurzivního neporozumění mezi chápáním tématu v řádu literárněvědné interpretace a mezi jazykem obsahové analýzy informačních věd pak vyplynulo i to, že první výsledky experimentů s různými metodami mohly být v závěru Bendíkovy práce na příkladu jedné konkrétní básně hodnoceny jako zdařilé, byť s tématem textu se z pohledu literární vědy poněkud minuly. Inkriminovanou básně, plnou motiviky námořních plaveb a v závěru pak i trosečnictví na opuštěném ostrově,<sup>11</sup> všechny použité metody v rámci supervizovaných, semisupervizovaných i nesupervizovaných

9 V informatice je definována jako možná „datová nouze“ svého druhu („data scarcity and sparsity“). Jde o to, že i v takto velkém datasetu mohou být jednotlivé texty vzájemně až příliš rozdílné, což pak činí obtíže při jejich seskupování do kategorií na základě podobnosti (BENDÍK 2023). Během značkování vzorových básní pro potřeby experimentu jsme si v rámci badatelského literárněvědného týmu nad konkrétním materiálem z různých dekád 19. století velmi dobře uvědomovali to, co reflektuje literární historie v oblasti vývoje poetiky: milostná básně puchmajerovců vypadá nejen formálními prostředky, ale celou výstavbou svého významu zcela jinak než verše na stejné téma z konce 19. století. Pro strojovou analýzu nakonec ovšem nemusí být rozpoznatelné, že jsou si takovéto texty svým obsahem dostatečně podobné.

10 Informatika pracuje v souvislosti se strojovou obsahovou analýzou s termínem *topics*, klíčových slov pro obsah textů. Pro přehled experimentů s analýzou topics na materiálu poetických textů různých jazykových oblastí viz PLECHÁČ — HAIDER 2020.

11 „Tvá loď jde po vysokém moři, | v ně brázdu jako stříbro reje, | svou přídu v modré vlny noří | a bok svůj pěnné do přeje. || Tvá lana sviští, plachty duní | a třepe vlnjaka. V noční chvíli | zříš magický svit mořských tůní, | a ve snu, Albatros jak pílí. ||

experimentů přiřadily k tematické oblasti „exotika/cestovatelství“. <sup>12</sup> Báseň napsal Eduard Albert a patří k těm, u nichž naneštěstí pro strojovou analýzu klíč k interpretaci tématu představuje v prvé řadě její titul — text nese název „Jaroslavu Vrchlickému“ a toto vodítko podporují rovněž další již ryze vñetextové informace o básni a jejím autorovi. <sup>13</sup> Veškerá zde použitá motivika z oblasti námořních cest a robinsonovských aluzí je tak jen působivou obraznou řečí o něčem jiném — konkrétně o vlastním vztahu autorského subjektu k Jaroslavu Vrchlickému; vyjadřuje se zde obdiv k velikánovi i zdrženlivé hodnocení vlastní tvorby ve srovnání s ním. <sup>14</sup>

Již v Bendíkově práci jsou co do koherence a přesnosti výsledků obsahové analýzy nejlépe vyhodnoceny metody nesupervizované (bez účasti předem anotovaných cvičebných dat), zvláště pak metoda Top2Vec (BENDÍK 2023: 67). Jak potvrdily i další experimenty, jež po Martinu Bendíkovi dále s využitím rozsáhlejšího datasetu anotovaných básní prováděli Karel Klouda a Magda Friedjungová, výsledky supervizovaných metod nejsou natolik přesvědčivé, aby z nich mohl být užitečný nástroj uživatelů ČEK, neboť skutečné téma básně tato analýza příliš často mívá (o úspěšném určení tématu bylo možné mluvit v méně než čtvrtině výsledků). Naproti tomu zaměření pozornosti na analýzu nižších významových jednotek obsahu básně (s jistým zjednodušením řečeno na analýzu motivickou) v rámci nesupervizovaných metod zpracování přineslo výsledky, jež se nástrojem databáze stát mohou, ba mohou se stát i užitečným nástrojem badatelským, budeme-li si vědomi jejich experimentální povahy a omezení, která z toho vyplývají.

Nesupervizované metody (tedy metody bez účasti lidského „učitele“) třídí básně do shluků (klastřů). Vzhledem k tomu, že v tomto případě nelze zkontrolovat přesnost výsledků na anotovaných datech, pracují tyto metody s tematickou koherencí a diverzitou textů (ABDELRAZEK 2023; ZHAO 2021):

---

Já samotním, jsem na ostrově, | ohýnek topím, rybkou lově | zasedám na břeh za večera. || Dým v kotoučích se modrých krade, | kdes písklo ptáče, ještě mladé, | tma na mne hrozí z pološera“ (ALBERT 1900: [XV]).

- <sup>12</sup> Martin Bendík pak zhodnotil téma básně následovně: „Tato báseň pojednává o scénách z námořní cesty, důraz je položen na lodě, vlny, plachty a noc na ostrově, vše je zahaleno ovzduším nostalgie a tajemna“ (IBID.: 63 [překlad z angličtiny LK]).
- <sup>13</sup> Eduard Albert byl sice povoláním chirurg a lékař, ale také velký propagátor a mecenáš české kultury a mezi jeho přátele náležel mj. Vrchlický. Albert Vrchlického obdivoval, napsal o něm monografii a také se právě jemu již ve zralém věku svěřil s vlastními veršovanými pokusy. Vrchlický je ocenil, připomínkoval a nakonec z nich sestavil sbírku *Na zemi a na nebi*, již zmiňovaná báseň otvírá dedikací právě Vrchlickému (sbírka vyšla roku 1900, autor se jí však nedočkal, nedlouho před jejím vydáním nečekaně zemřel).
- <sup>14</sup> Tuto linku podporují i formální charakteristiky — báseň je italským sonetem v jambu, tedy formou Jaroslavem Vrchlickým zvláště oblíbenou.

do klastru se shlukují básně, jež jsou si vzájemně obsahově co nejpodobnější, a mezi jednotlivými klastry by naopak měly být rozdíly co největší. Cílem experimentů bylo vytvořit všechny klastry řádově stejně velké; přičemž každá báseň je nakonec zařazena právě do jednoho klastru; duplicitní znění týchž textů byla pro výpočet odstraněna (pracovalo se jen s jejich prvními vydáními).

Jako nejúspěšnější pro účely projektu se ukázal model Top2Vec s vnořením vytvořeným pomocí doc2vec modelu (ANGELOV 2020). Tato metoda vytváří vektorovou reprezentaci nejen pro každou báseň, ale i pro její jednotlivé tokeny (autosémantická lemmatizovaná slova). Každý klaster je posléze charakterizován vždy výčtem deseti lemmatizovaných autosemantických lexikálních jednotek, jež se v textech daného klastru vyskytují s největší pravděpodobností, neboť daná metoda umí slova reprezentující daný shluk najít jako ta, která jsou danému shluku nejbližší s tím, že tato vzdálenost se počítá právě pomocí blízkosti těchto vektorových reprezentací. Vektorovou vzdálenost konkrétní básně od pomyslného centra klastru charakterizuje její skóre — čím vyšší je, tím podobnější sémantickému jádru konkrétního shluku by daná báseň měla být. Pro potřeby přehledného zobrazení klastrů uživatelům aplikace ČEK jsou básně daného klastru seřazeny od nejvyššího skóre po nejnižší.

Na výsledcích takového zpracování dat může být pro literární vědu zajímavé především to, že výpočetní technika zvládne při strojové analýze v relativně únosném čase nižších desítek hodin zpracovat veškerý básnický materiál z databáze a dosáhnout jeho strukturace, která vzhledem k rozsahu těchto dat není v možnostech individuální lidské mysli. Skrze tuto strukturu se pak do tohoto celku nabízí nový vhled, který může otevírat i cestu k novým interpretacím. V další části tohoto textu upozorníme na několik konkrétních příkladů výsledků této analýzy a pokusíme se naznačit směry možného badatelského využití získaných výsledků.

Při zpracování ČEK byla její data strukturována do 106 klastrů (počet klastrů si model Top2Vec určuje po nastavení základních výchozích parametrů vždy zcela automaticky sám). Klíčová slova charakterizující každý klaster (může jít o apelativa i propria) mohou navést naši pozornost k motivice, která je dostatečně unikátní v celku dat a zároveň je v něm dostatečně frekventovaně zastoupená.

Výskyt některých klastrů historika znalého kulturně-společenského dění 19. století spíše nepřekvapí. Hned dva z klastrů tak např. zcela zjevně dokládají afinitu českého kulturního prostředí k slovenskému národu a jeho prostoru, spojenou také s vyhraňováním se vůči politikám maďarských elit (první klaster charakterizují klíčová slova *Slovač, Slovensko, slovenský, Tatry, Kriváň, valaška, šuhaj, koliba, Slovák, Váh*; druhý pak klíčová slova *Slovák, Maďar, Slovač, slovenský, maďarský, tatranský, Slovensko, Tatry, Kriváň, Tatra*). Jiný klaster

i jen skrze svá klíčová slova (*vlak, nádraží, kolejnice, kupé, stanice, kolej, rychlík, vagón, tunel, hvízdá*) vypovídá o zásadní transformační roli rozšíření vlakové dopravy v druhé půli 19. století a o dobové fascinaci vlakem. Zároveň platí, že konkrétní texty zařazené do příslušného klastru jistě otevírají další možnosti, jak se tomu kterému společenskému fenoménu lze věnovat — například lze sledovat proměny převládajících přístupů k danému tématu v běhu času či v různých ideových rámcích.

To stejné lze také činit nad materiálem klastrů, které se vyprofilovaly mimo samozřejmá očekávání. Ukazuje se například, že česká poezie dlouhého 19. století obsahuje dostatek textů s výraznou motivikou Benátek (klíčová slova klastru *laguna, gondola, Benátky, signorina, kanál, loďka, Lido, kráska, člun, mramorozý*), ale také třeba samostatný klastr zahrnující ukrajinské realie (klíčová slova *ataman, kozák, síč, step, kozácký, Ukrajina, Dněpr, junák, Turek, Kyjev*), nebo abychom nezůstávali jen v rovině místopisu, klastr s motivikou loveckou (klíčová slova *jelen, lovec, paroh, laň, srna, pylák, srnec, myslivec, lovcí, zvěř*) či třeba mlynářskou (klíčová slova *mleč, mlynář, mlýn, stárek, mlejnice, krajánek, mlít, mlynářka, stavídlo, mlýnice*). Některé klastry zase překvapí svým autorským obsazením — hned ten vůbec nejpočetnější (čítající i 658 básní), charakterizovaný klíčovými slovy *viditelný, mystický, tajemný, tajemství, vegetace, tisíciletí, mlčení, magický, závat, agonie*, je obsazen na řadě prvních pozic dle skóre výhradně texty Otokara Březiny (teprve na 33. pozici v klastru se vyskytne první báseň od jiného autora/autorky, konkrétně text Růženy Jesenské, Březinovo autorství je ovšem téměř výhradní i na dalších pozicích a alespoň poněkud se začne tento klastr autorsky diferencovat teprve v sedmé desítkce sem spadajících básní). Strojová analýza tak jen potvrzuje silnou pozici autorské motiviky, jež je v celku dat dostatečně frekventovaně zastoupena, což zjevně znamená i to, že se od sbírky ke sbírce výrazněji neproměnila.

Za překvapivý na první pohled lze prohlásit i co do počtu zařazených básní pátý nejrozsáhlejší klastr, odkazující k lexiku spjatému s mořem a mořeplavbou, jak jej vystihují klíčová slova *loď, plachta, plavec, stožár, člun, přístav, koráb, paluba, plout, vrak*. Podíváme-li se nicméně blíže na texty, jejichž vzdálenost je od jádra klastru nejmenší (tj. takové, u nichž je sémantická přínaléžítost k dané skupině největší vůbec), rychle zjistíme, že z největší části nevyjadřují skrytou touhu české společnosti odkázané k životu ve vnitrozemí po mořském pobřeží. Plavba po moři se v těchto textech stává rozvitou metaforou běhu života, jak napoví již podtitul v pořadí druhé básně klastru „Život lidský — plavba širým mořem“ (SUŠIL 1869). Téma básní využívajících tuto motiviku ovšem může být velmi různorodé. U nábožensky orientovaných autorů, jejichž texty se dostaly do předních pozic tohoto klastru, je konečným cílem metaforické plavby až spása duše v království nebeském, jedině Bůh může být v nevyzpytatelnosti mořských bouří (alias tragických životních událostí) bezpečným navigátorem (viz odkazovanou báseň katolického kněze Františka



Sušila [SUŠIL 1869]) a je třeba se k němu obrátit dřív, než dojde ke ztroskotání („patřte k nebi, pokud ještě jasně: | pozdě již, kdy loď se troskotá!“, jak radí jiný katolický kněz Jan Ježek [JEŽEK 1880]). Záchranu z širého moře životních peripetií, zbavenou všech náboženských aluzí, může také představovat intimní vztah motivicky převlečený za „maják záchranný“ — jako ve stejnojmenné básni Rudolfa Illového (ILLOVÝ 1908). Moře může být vnímáno jako prostor nebezpečí a osamělosti těch, kdo se na ně vydávají, a to v zásadě v realisticky pojatých obrazech (SLÁDEK 1881), ale stejně tak může být tento prostor využit k metaforickému pojednání životní osamělosti (KAMINSKÝ 1893) či nebezpečných životních ztroskotání (QUIS 1872); třeba u S. K. Neumanna je moře obrazně pojato coby prostor sice nutně nebezpečné, ale zároveň opojné individuální svobody (NEUMANN 1903). Výše zmiňovaná básně Eduarda Alberta, využívající motiviku moře a námořních cest k tematizaci literární tvorby, je v klastru rovněž přítomna, a to relativně blízko k jeho centru, na 46. pozici dle skóre. Motivika je tu napříč básněmi shodná; tematicky se sice mezi nimi nalézou i texty, které jsou si relativně blízké, celkově je ale rozptýl témat značný.

V některých případech nicméně motivický klastr může naši pozornost naměřovat i k relativně užšímu tematickému okruhu. Z tohoto hlediska se nám jeví jako zvláště nosný zejména klastr charakterizovaný slovy *nit*, *len*, *příst*, *kolovrátek*, *cívka*, *přádlo*, *kužel*, *příze*, *niika*, *vrčet*, a to zejména proto, že v něm nacházíme s velkým zastoupením tematizaci přádlu coby typicky ženské práce,<sup>15</sup> která se zároveň těsně provazuje s něčím, co lze snad nazvat představou o „ženském údělu“ a „ženských ctnostech“. Dobrá žena je ta, která skromně a pilně přede kdesi v ústraní (srov. PICEK 1856), kde si ji případně vyhlíží její nastávající životní partner (HEYDUK 1888; ERBEN 1853a, 1853b); a dobrá žena je ta, která zůstává stejně pokorná při výrobě textilií do své výbavy (HEYDUK 1888), textilií na zaopatření svých dětí (v krajním případě i ze záhrobí [IDEM 1899]) i při chytání přádlu na rubáš svým blízkým, a nakonec i sama sobě (NERUDA 1868). V básních tohoto klastru se přástky objevují jako zvláštní čas umožňující vesnické mládeži družbu a namlouvání (PROCHÁZKA 1899; POKORNÝ 1887; ŠNAJDR 1830); na druhou stranu se výroba přediva, případně textilií z něj celkem pravidelně stává metaforou odvíjení běhu života.

Pro tento i pro všechny ostatní klastry nicméně platí: je třeba mít za všech okolností na paměti, že jde o výsledky strojového zpracování. I proto je nutné zkoumat takřka výhradně ty texty, jež jsou shluknuty relativně blízko centra daného klastru. Básně vyskytující se na periferii klastrů již žádnou zásadní souvislost s daným sémantickým polem mít nemusejí, respektive k jejich za-

<sup>15</sup> Naproti tomu tkaní se častěji pojednává v rámci sociálně zaměřeného tematizování práce chudých tkalců-mužů — srov. do klastru zařazené básně Mužikovy (1883, 1902), případně celý cyklus „Tkalcovské písně“ Josefa Kaluse (1895), z nějž se ovšem do klastru neshlukly všechny básně.

řazení právě do tohoto shluku mohl vést například výskyt podobného lexika, které je v konkrétních textech ovšem použito v posunutém významu.<sup>16</sup>

Napříč daty se bude rovněž lišit badatelská užitečnost jednotlivých klastrů. Je třeba počítat i s možností, že některé jsou spíše dílem nevysledovatelných souvislostí, případně mohou do výsledku promlouvat třeba frekventovaně užívané rýmy některých slov. Posledně zmíněná možnost je velmi pravděpodobným vysvětlením klasteru s klíčovými slovy, jež jsou až na jedinou výjimku samými deminutivy: *Hanička, Jeníček, hubička, růžička, Hanna, panenka, srděčko, srdíčko, děvčátko, Janíček*. Možnost rýmovat domáckou podobu jména *Hanička* se zdobnělymi apelativními femininy a s genitivem a akuzativem zdobnělého jména *Jeníček/Janíček*, vokativní *Haničko* pak i s deminutivy středního rodu vzoru *město* zjevně lákala zejména tvůrce tíhnoucí k idyle. V pořadí druhá báseň klasteru, „Pomněnky“ Františka Jaroslava Kamenického (1833), zaujme schopností v duchu *biedermeieru* použitými deminutivy ohladit tragické hrany života: v prvních strofách vše krásné připomíná lyrickému subjektu jeho milou Haničku,<sup>17</sup> teprve postupně se čtenář dozvídá, že Hanička již není mezi živými;<sup>18</sup> útěšnou perspektivu pak poskytuje mluvčímu víra ve zmrtvýchvstání — i ta je nicméně vyjádřená s využitím zdobnělín.<sup>19</sup> Česká poezie dokázala skutečnost, že rýmování s využitím deminutiv a domácké podoby jména *Hanička* je v ní přítomno s relativně vysokou frekvencí, sama reflektovat — a tato reflexe je dokonce přítomna i v dotyčném klasteru. Na prvním místě, tedy coby nejbližší ryzí realizace sémantiky jeho klíčových slov, tak shluku vévodí text „Černooká Hanička“ Jana Duchoslava Panýrka (1862), který právě sentimentální veršování s využitím patřičných deminutiv paroduje:

Černooká Hanička  
milovala Jeníčka;  
Jeníček zas Haničku  
nosil ve svém srdíčku.

16 Máme-li zůstat u příkladu periferie stejného klasteru, jeho předposlední básní je text „Sestra Nina“ od Ferdinanda Píseckého (1919), který pojednává o tom, jak titulní hrdinka staršího věku opouští naději na vlastní milostné štěstí a oddává se obětavé práci v lazaretu. Lexikální souvislosti, která vedla k zařazení právě sem, jsou tu ovšem patrně jen metaforické „nitky“ jejích stříbrných vlasů, signalizující odcházející mládí.

17 „Vidím-li v naší zahrádce | Z jara růžičky, | Myslím si: Takové měla | Hanna tvářičky. || [...] V zimě když bělavý lníček | Předě maticka, | Zpomínám: Takové vlasy | Měla Hanička“ [...].

18 „Když v tom našem kostelíčku | Vidím anděla, | Myslím si: Dušička její | K nebi letěla“ [...].

19 „A když nad hřbitovem naším | Vstává sluníčko, | Zvolám: Někdy z hrobu vstaneš | I ty Haničko!“

Plyne čas co vodička —  
nevěrného Jeníčka  
oplakává Hanička —  
je to stará písnička.

Strojová analýza ironii přehnaného použití typického lexika odhalit neumí; stejně jako výše zmiňované zohlednění metadat, intertextových souvislostí a širšího literárněhistorického kontextu toto zůstává i nadále čistě lidským interpretačním výkonem. Strojová analýza nicméně umí, a to poměrně efektivně, k takovým interpretačním výkonům nachystat svým tříděním možný materiál.

Text vznikl v rámci grantového projektu TA ČR L05000288  
„Analýza motivických klastrů z oblasti aktuálních kulturně-společenských  
témat a jejich aplikace na materiál uměleckých textů 19. a počátku 20. století“  
řešeného v Ústavu pro českou literaturu AV ČR, v. v. i.

Při práci na textu byly využity zdroje výzkumné infrastruktury  
Česká literární bibliografie — <https://clb.ucl.cas.cz/> (kód ORJ: 90243).

### Prameny

(všechny níže uvedené básně odkazujeme k jejich prvnímu knižnímu vydání, jsou nicméně pochopitelně přítomny v databázi Česká elektronická knihovna, kde jsme s nimi pracovali)

ALBERT, Eduard

**1900** „Jaroslavu Vrchlickému“; in idem: *Na zemi a na nebi* (Praha: František Šimáček), s. [XV]

ERBEN, Karel Jaromír

**1853a** „Zlatý kolovrat“; in idem: *Kytice z pověstí národních* (Praha: Jaroslav Pospíšil), s. 42–57

**1853b** „Štědrý den“; in idem: *Kytice z pověstí národních* (Praha: Jaroslav Pospíšil), s. 58–66

HEYDUK, Adolf

**1888** „Přadlena“; in idem: *Obrázky* (Praha: Eduard Valečka), s. 38–42

**1899** „Zlaté přediwo“; in idem: *Za dlouhých večerů* (Praha: Jan Otto), s. 76–80

ILLOVÝ, Rudolf

**1908** „Maják záchranný“; in idem: *Kročjeje* (Praha: Otakar Janáček), s. 52

JEŽEK, Jan

**1880** „Naše žití“; in idem: *Za šera* (Praha: nákladem vlastním), s. 76

KALUS, Josef

**1895** „Tkalcovské písně“ [cyklus]; in idem: *Kresby a písně* (Praha: Jan Otto), s. 29–46

KAMENICKÝ, František Jaroslav

1833 „7. Pomněnky“; in idem: *Písně v národním českém duchu* (Praha: Václav Špinka), s. 20–22

KAMINSKÝ, Bohdan

1893 „Bez přístavu“; in idem: *V samotách* (Praha: Jan Otto), s. 30

MUŽÍK, Augustin Eugen

1883 „Píseň tkalce“; in idem: *Hlasy člověka* (Praha: Josef Richard Vilímek), s. 104–107

1902 „Tkalci“; in idem: *Krůpěje rosy* (Praha: Jan Otto), s. 102

NERUDA, Jan

1868 „Kolovrátek“; in idem: *Knihy veršů* (Praha: Eduard Grégr), s. 31–39

NEUMANN, Stanislav Kostka

1903 „Námořnická písnička“; in idem: *Sen o zástupu zoufajících a jiné básně* (Praha: Moderní revue), s. 28

PANÝREK, Jan Duchoslav

1862 „XIV. Černooká Hanička“; in idem: *Písně z hladu* (Praha: Jaroslav Pospíšil), s. 16

PICEK, Václav Jaromír

1856 „Nevěsta předoucí“; in idem: *Písně 4* (Praha: nákladem vlastním), s. II

PÍSECKÝ, Ferdinand

1919 „Sestra Nina“; in idem: *Když tekla krev* (Praha: Bedřich Kočí), s. 47–48

POKORNÝ, Rudolf

1887 „Zima v horách“; in idem: *Opět na horách* (Praha: Eduard Valečka), s. 60–62

POLOM, J. B.

1910 „Stařenčin odkaz“; in idem: *Balady a dojmy slovenské* (Praha: Jaroslav Bursík/František Kohout), s. 24–25

PROCHÁZKA, František Serafinský

1899 „XIII. Přádka“; in idem: *Novější písničky* (Praha: Josef Richard Vilímek), s. 30–31

QUIS, Ladislav

1872 „V bouři života“; in idem: *Z ruchu* (Praha: nákladem vydavatelovým), s. 121–122

SLÁDEK, Josef Václav

1881 „Stíny mrtvých“; in idem: *Světlou stopou* (Praha: Lumír), s. 134–135

SUŠIL, František

1869 „80. Život lidský — plavba širým mořem“; in idem: *Zpěvy a hněvy* (Brno: Matice moravská), s. 84

ŠNAJDR, Karel Sudomír

1830 „Kolovrátek“; in idem: *Okus v básněni českém 2* (Hradec Králové: Jan Hostivít Pospíšil), s. 28–30

## Literatura

ABDELRAZEK, Aly et al.

- 2023 „Topic modeling algorithms and applications: A survey“; *Information Systems*. February 2023, 112(C), pp. 102–131. [cit. 2023-02-14]; dostupné na: <https://doi.org/10.1016/j.is.2022.102131>

ANGELOV, Dimitar

- 2020 „Top2Vec: Distributed Representations of Topics“; arXiv preprint arXiv:2008.09470. August 2020. [cit. 2023-02-15]; dostupné na: <https://arxiv.org/pdf/2008.09470.pdf>

BENDÍK, Martin

- 2023 *Automatic detection of topics in poetic texts* [diplomová práce; obhájeno na Fakultě informačních technologií ČVUT, Katedře aplikované matematiky, vedoucí práce: Karel Klouda]; dostupné na: <https://dspace.cvut.cz/handle/10467/108779>

FOŘT, Bohumil

- 2018 „Téma“; in Ondřej Sládek (ed.): *Slovník literárněvědného strukturalismu* (Praha/Brno: ÚČL AV ČR/Host), s. 751–754

KOLÁR, Robert

- 2023 *Hudbu vašich rytmě jsem cítil...: průvodce veršovanými prostředky mezitextového nazarování v díle V. Dyka, F. Gellnera, F. Šrámka a K. Tomana* (Praha: Akropolis/ÚČL AV ČR)

LUBKOLL, Christine

- 2006 „Motiv, literární“; in Ansgar Nünning (ed.): *Lexikon teorie literatury a kultury* (Brno: Host), s. 530–531

MACURA, Vladimír

- 1999 *Český sen* (Praha: Nakladatelství Lidové noviny)

O PROJEKTU

- 2005 [úvodní text na webové stránce České elektronické knihovny]; <https://www.ceska-poezie.cz/>, [cit. 9. 1. 2023]

PLECHÁČ, Petr — HAIDER, Thomas N.

- 2020 „Mapping topic evolution across poetic traditions“; arXiv preprint arXiv:2006.15732. June 2020. [cit. 2023-02-14]; dostupné na: <https://arxiv.org/pdf/2006.15732.pdf>

SLÁDEK, Ondřej — FOŘT, Bohumil

- 2018 „Motiv“; in Ondřej Sládek (ed.): *Slovník literárněvědného strukturalismu* (Praha/Brno: ÚČL AV ČR/Host), s. 472–477

SVADBOVÁ, Blanka

- 2000 „Plnotextová databáze české poezie 19. století“; *Česká literatura* XLVIII, č. 3, s. 318–328

ŠIDÁK, Pavel

- 2012 „Motiv“; in Richard Müller, Pavel Šidák (edd.): *Slovník novější literární teorie* (Praha: Academia), s. 327–330

ZHAO, He et al.

2021 „Topic modelling meets deep neural networks: A survey“; arXiv preprint arXiv:2103.00498. February 2021. [cit. 2023-02-15]; dostupné na: <https://arxiv.org/pdf/2103.00498.pdf>

---

### Creative Commons

Toto dílo podléhá licenci Creative Commons — Uvedte původ — Neužívejte komerčně — Nezpracovávejte 4.0 Mezinárodní licence. Licenční podmínky jsou dostupné zde: <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode.cs>.

