

# Pokroky matematiky, fyziky a astronomie

---

Mojmír Otruba; Pavel Vašák

První užití matematiky v české jazykovědě

*Pokroky matematiky, fyziky a astronomie*, Vol. 19 (1974), No. 2, 86--91

Persistent URL: <http://dml.cz/dmlcz/139236>

## Terms of use:

© Jednota českých matematiků a fyziků, 1974

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* <http://project.dml.cz>

# První užití matematiky v české jazykovědě

*Mojmír Otruba, Pavel Vašák, Praha*

Domněle staročeské památky Rukopis královédvorský (RK) a Rukopis zelenohorský (RZ), objevené v l. 1817–18, byly českou vědou přijaty jako díla 13. a 9.–10. století. Jediný J. DOBROVSKÝ, který sice přijal RK, odmítl RZ hned po jeho objevení jako falzum. Česká věda na pravosti RKZ více než půlstoletí důsledně trvala a jednoznačně zamítala i velmi seriózní námítky proti jejich pravosti pronášené zahraničními odborníky (zvl. B. KOPITAREM, J. FEIFALIKEM, V. JAGÍCEM). Představa, že RKZ jsou díla pravá, stala se v 19. stol. také součástí obecného kulturního povědomí u nás, stejně jako názor, že námítky proti pravosti jsou veskrze motivovány nacionalistickou averzí. Tato situace umožňuje pochopit, že tzv. rukopisné spory se staly doslova celonárodní kulturní záležitostí v okamžiku, kdy se proti pravosti začínali vyslovovat nikoli již zahraniční, nýbrž čeští vědci. Po vystoupení jednotlivců, A. V. ŠEMBERY (1878) a A. VAŠKA (1879), dala GEBAUEROVA stať *Potřeba dalších zkoušek Rukopisu královédvorského a zelenohorského* (Athenaeum 3, 1886, 152–165) podnět na jedné straně k odborné revizi RKZ, na druhé straně k vehementní obraně pravosti, v níž více mluvila žurnalistika než odbornost.

Z hlediska metodického rozvoje české vědy je závažné, že se na zkoumání pravosti RKZ, soustředěného v MASARYKOVĚ revui Athenaeum, podílela spolu s jazykovědou i politická a literární historie, sociologie, estetika, paleografie aj., a to v takové návaznosti věcné i metodicky programové, že lze tento průzkum označit jako komplexní. K disciplínám, do jejichž oblasti zkoumání náležely RKZ přímo (neboť byly fenoménem literárním, jazykovým, historickodokumentárním aj.), přistoupily i další obory, které se nemohly vyjadřovat k samotnému obsahu sporu, ale mohly mít vliv na objasnění problémů a vůbec na myšlení odborníků, a to uplatňováním metodologických zkušeností. Takto se na komplexním průzkumu RKZ podílela textová kritika klasické a germánské filologie (J. KRÁL a A. KRAUS), takto také poprvé u nás vstoupila svými poznávacími možnostmi do problematiky humanitních věd matematika.

Profesor AUGUSTIN SEYDLER – který byl od prvního ročníku v Athenaeu odborným redaktorem pro filosofii a již touto svou redakční činností se dostával do styku se sporem o RKZ – odborně přispěl k diskutované záležitosti svou statí *Počet pravděpodobnosti v přítomném sporu* (Athenaeum 3, 1886, 299–307). Článek byl publikován dva měsíce po zmíněné studii Gebauerově a po vypuknutí sporu, tedy v době, kdy již kritika upozornila na takové množství sporných jevů v jazyce RKZ, že se mohl klást problém zákonitosti a nahodilosti tohoto úkazu, tj. problém matematicky formulovatelný a řešitelný.

A. Seydler sáhl po počtu pravděpodobnosti; materiálem, který měl být takto vyhodnocen, jsou mu Gebauerova zjištění o RK. J. Gebauer, který vycházel z představy „mluvnice normální“, budované na obvyklých a doložitelných jazykových jevech (hláskových, tvarových, skladebných) v prokazatelně pravých staročeských památkách, zjistil v jazyce

RK velké kvantum odchylek. Seydler v návaznosti na zjištění Gebauerova zavádí tu axiologicky neutrální pojem *zvláštnosti*. Část těchto zvláštností je navíc pozoruhodná tím, že nemá oporu v staročeských jazykových projevech, má však totožný nebo analogický doklad buď v prokázaných novodobých padělcích jiných, nebo v chybných odborných výkladech a aplikacích staré češtiny pocházejících z doby nedlouho před objevením RKZ, nebo v dílech novočeské beletrie rovněž z doby před objevením RKZ.

Seydler se v dalších uvahách pro zjednodušení omezuje pouze na RK. Shrňme jeho výchozí předpoklady, podle kterých jsou zvláštnosti vyskytující se v RK dvojího druhu:

- (1) zvláštnosti vyskytující se pouze v RK;
- (2) zvláštnosti vyskytující se zároveň v RK a v některých jiných padělaných spisech pocházejících z 19. stol., z doby těsně před objevením RK.

Seydler chce proto zjistit, zda je původ těchto zvláštností náhodný a zda je náhodná souvislost RK a textů prokazatelně padělaných. Pokusíme se vyložit, jak Seydler postupoval.\*)

K vysvětlení původu všech zvláštností jsou obecně možné různé hypotézy. Předpokládejme, že tyto hypotézy tvoří konečnou množinu a označme ji jako  $\mathcal{H}$ . Uveďme některé důležité prvky této množiny:

- (1) Hypotéza náhodnosti, tj. zvláštnosti vznikly náhodně, např. jako roztržitost a omyly písařů, opisovačů aj., vlivem času porušené písmo atd. Hypotézu označíme symbolem  $H_N$ .
- (2) Hypotéza jazykové zákonitosti, tj. uvažované zvláštnosti nejsou zvláštnostmi v pravém slova smyslu, ale výrazem nám utajené jazykové zákonitosti, např. neznámý staročeský dialekt, básnický jazyk apod. Tuto hypotézu označíme jako  $H_Z$ .

Podoba zbývajících hypotéz z množiny  $\mathcal{H}$  se při Seydlerově elementárním postupu neuvažuje.\*\*)

Uvažujme nyní nějaký gramatický jev v RK. Budiž  $c$  celkový počet jeho výskytů v celém textu a  $0 \leq a \leq c$  nechť označuje počet zvláštností (např. v souladu s Gebauerem z celkového počtu  $c = 89$  imperfekt v RK je jich  $a = 17$  odchýlných). Počet zvláštností  $a$  můžeme považovat za realizaci náhodné veličiny  $Z$ , která nabývá celých hodnot

$$Z = z, \quad 0 \leq z \leq c$$

a jejíž pravděpodobnostní rozdělení

$$P(Z = z|H), \quad 0 \leq z \leq c$$

\*) Za pomoc při rozboru ne vždy přesných a úplných Seydlerových úvah autoři děkují dr. J. Štěpánovi z MFF KU.

\*\*) Při výkladu pojmů pravděpodobnost, hypotéza aj. Seydler čtenáře odkazuje na čl. A. PÁNKA, *Pravděpodobnost a posteriori*, Časopis pro pěstování matematiky a fyziky XII, 1883, 227–232; dále na práci S. D. POISSON, *Lehrbuch der Wahrscheinlichkeitsrechnung*, německý překlad od SCHNUSSE z r. 1841.

závisí ovšem na tvaru zvolené hypotézy  $H \in \mathcal{H}$ . Je-li např.  $H = H_Z$ , pak

$$(1) \quad \begin{aligned} P(Z = a|H_Z) &= 1, \\ P(Z = z|H_Z) &= 0 \quad \text{pro } z \neq a. \end{aligned}$$

Jestliže přijmeme hypotézu nám utajené jazykové zákonitosti, je počet tzv. zvláštností determinován naší neznalostí gramatické struktury staré češtiny, tj. v podstatě jejím neúplným popisem (dialekt, básnický jazyk apod.). Naopak, položíme-li  $H = H_N$ , je

$$(2) \quad P(Z = z|H_N) = \binom{c}{z} x^z (1-x)^{c-z}, \quad 0 \leq z \leq c.$$

V každém z  $c$  případů můžeme totiž očekávat zvláštnost s neznámou pravděpodobností  $x$ , dobrý (správný) tvar s pravděpodobností  $(1-x)$ ; veličiny jsou nezávislé, jejich součet dá binomické rozložení. Neznámá pravděpodobnost  $0 \leq x \leq 1$  tedy udává pravděpodobnost písařských a jiných omylů různého druhu.

Skutečnost, že  $x$  je neznámé, řeší Seydler tím, že nachází maximum výrazu (2) při  $z = a$ .

Tedy

$$(3) \quad P(Z = a|H_N) \leq \binom{c}{a} \left(\frac{a}{c}\right)^a \left(\frac{c-a}{c}\right)^{c-a}.$$

Zajímá nás však aposteriorní rozdělení hypotézy  $H \in \mathcal{H}$  při skutečně pozorovaném počtu zvláštností  $Z = a$ , tj.

$$P(H|Z = a), \quad H \in \mathcal{H}.$$

Seydler se spokojuje s nalezením horního odhadu pro aposteriorní pravděpodobnost hypotézy náhodnosti při výskytu  $a$  odchylek

$$P(H_N|Z = a).$$

Poznamenejme, že Seydler – korektně a vědom si složitosti sporu – uvádí, že „čím je větší pravděpodobnost na straně pouhé náhody“, tím více to svědčí „ve prospěch pravosti Rukopisů“, podobně čím je větší pravděpodobnost „na straně zákonitosti oněch zvláštností a příčinného svazku vzhledem ke koincencím, tím více to svědčí proti pravosti, poněvadž tím více vzrůstají rozpaky, jak sobě onu zákonitost a tento příčinný svazek bez poškození víry v pravost upravit a představit“.

K přechodu od pravděpodobností  $P(Z = z|H)$  k pravděpodobnostem aposteriorním se užívá známý BAYESŮV vzorec

$$(4) \quad P(H_0|Z = z) = \frac{P(H_0) \cdot P(Z = z|H_0)}{\sum_{H \in \mathcal{H}} P(H) \cdot P(Z = z|H)} \quad \text{pro } H_0 \in \mathcal{H}.$$

K tomu je však třeba znát apriorní pravděpodobnosti jednotlivých hypotéz, tj.

$$P(H), \quad 0 \leq P(H) \leq 1, \quad H \in \mathcal{H}, \quad \sum_{H \in \mathcal{H}} P(H) = 1.$$

I když to Seydler přímo neuvádí, považuje pravděpodobnosti  $P(H)$  jednotlivých hypotéz za stejné, tj.

$$(5) \quad P(H) = \frac{1}{\text{card } \mathcal{H}} \quad \text{pro } H \in \mathcal{H} .$$

Staví se tak na pozici nestranného pozorovatele a nepreferuje žádnou z hypotéz. Celkem tedy ze vztahů (1), (4) a (5) vyplývá, že

$$P(H/Z = a) \leq P(Z = a/H) \quad \text{pro } H \in \mathcal{H} .$$

Na základě vztahu (3) dostáváme hledaný odhad pro aposteriorní pravděpodobnost hypotézy náhodnosti

$$P(H_N/Z = a) \leq \binom{c}{a} \left(\frac{a}{c}\right)^a \left(\frac{c-a}{c}\right)^{c-a} .$$

Na tento vztah aplikuje Gebauerovy výsledky: Podle Gebauera je v RK asi 5400 slov, z toho 700 tvarů je odchylných (zvláštností); Seydler z výpočtářského hlediska jich uvažuje pouze 540, tzn. že v každém desátém slově je odchylka od pravých staročeských památek. V RK je např. 89 imperfekt, mezi nimi je 17 odchylných tvarů; výpočtem Seydler ukazuje, že pravděpodobnost jejich náhodného výskytu je menší než  $1/9$ . Takových odchylných jazykových skupin je ovšem v RK mnohem víc (10); po rovnoměrném rozdělení slov i odchylek se každá skupina skládá z 540 slov, ve kterých je 54 odchylek. Podle Seydlera „zjistě tedy neukřivdíme hypotesi pouhé náhody, když jí připustíme za pravděpodobnost číslo shora nalezené, tj.  $1/9$ “. Potom je pravděpodobnost náhodného výskytu deseti odchylných skupin menší než  $1/9^{10}$ , tj.  $1 : 348 \cdot 10^7$ .

Podobným způsobem Seydler zavádí koeficient koincidence jazykových zvláštností mezi RK a staročeskými padělkami. Rovněž i zde ukazuje velkou nepravděpodobnost náhodné shody odchylných tvarů v RK a ve „staročeských“ textech HANKOVÝCH a jiných z doby před objevením RK.

A. Seydler si byl vědom toho, že svým způsobem vyhodnocování nivelizuje „zvláštnosti“ zjištěné Gebauerem, zatímco mezi těmito zvláštnostmi existují diference co do jazykovědné výpovědní hodnoty pro padělanost RK. Byl si vědom i toho, že správnost jeho dedukcí je vůbec nejprve závislá na správnosti poznatků Gebauerových, které přijímá jako pravdivé výroky, a dále teprve na správnosti matematických úvah. Vědomí, že matematikovo vyjádření k jazykovědné problematice je novum, jakož i znalost žurnalistických polemických zvyklostí tehdejší doby vyvolaly jeho vtipnou ironickou poznámku: „A tu již osměluji se prositi ne—odborníky, tj. ne—matematiky, aby svého práva blamovati se námitkami nedorozuměním vzniklými příliš nezneužívali.“

Na Seydlerovo vystoupení v Athenaeu reagoval v Národních listech bombastický obhájce RKZ dr. J. GRÉGR jen slovním znevažováním: „K řadě dosavadních útočnickův přistoupili útočníci noví; zorganizovala se hotová literární mafia, která nesahá sice do kapes jednotlivcův, ale vztahuje své smělé ruce po nejdražších skvostech národa. Je to kotérie na mladé naší universitě, která opatřila sobě již i svého šaška, jako jej mívaly

středověké „rytířské“ výpravy, když šlo o to, zaskočít a obrat kupce, vracející se s vzácným zbožím z dalekého trhu.“ Jak z dalšího vyplývá, tímto „šaškem“ je označován „šprýmovný matematik“ Seydler; Grégr mu připomíná latinské úsloví „si tacuisses, philosophus mansisses“, srovnává ho s „důmyslnými badateli středověku“, počítajícími „kolik miliónů čertů může tančit na špičce jehly“, lituje ho, že „s takovou při mužích pracujících ve vědě řídkou obětavostí přinesl ve sporu o naše rukopisy svému soudruhu T. G. Masarykovi za oběť i svou vědeckou reputaci“, atd. „Když již jde se na naše rukopisy s takovými to matematickými formulami, můžeme se toho nadíti, že v nejbližší příštím čísle Athenaea rozmnoží se společnost našich ‚vědců‘ ještě i nějakým spiritistou, který bude citovat proti našim rukopisům duchy...“ (Knižně in: J. GRÉGR, *Na obranu RKZ*, 1886, 74–77).

Historik prof. J. KALOUSEK jako jediný z obhájců zřejmě porozuměl Seydlerovým matematickým dedukcím a pochopil i jejich argumentační polemickou sílu. Ve své odpovědi (*Osvěta* 16, 1886, 556–558) nereplikuje však ani on na vlastní matematický problém, nýbrž na to, jak pravdivá jsou jazyková zjištění zvláštností, jež jsou matematicky vyhodnocována: „...o matematické dedukce ani nezavadím, není mně toho potřeba. Stojím před nimi jako kupčíkladu před mlýnem, jehož strojům nemusím rozuměti. Já jen dávám pozor, co se do mlýna vozí a na koš sype...“ Ze Seydlerovy odpovědi Kalouskovi (*Dodatek k mé úvaze o pravděpodobnosti*, Athenaeum 3, 1886, 446–448) je – kromě odkazu na případ vědeckého sporu o tzv. dvojhvězdy, v němž J. MITCHELL r. 1717 právě pomocí pravděpodobnosti podepřel názor o skutečném, nikoli optickém sousedství dvou hvězd v dvojhvězdách – nejzávažnější Seydlerovo konstatování: „Nechci dokazovati matematikou něco, co může dokázati jen lingvistika, paleografie nebo historie (sociologie); vida však, jak málo se cítí váha pochybností těmito vědami, zejména první z nich, pronešených, chtěl jsem, pokud to bylo lze, váhu tu číselně vyjádřiti.“

Matematika tedy vstoupila v tomto prvním případě do jazykovědy více jako prostředek zesilující hlas lingvistiky v polemické vřavě než jako nástroj napomáhající jazykovědnému poznání. Tímto konstatováním nemá být snižen historický význam Seydlerova užití matematiky v jazykovědě; spíše je třeba litovat toho, že vzrušené období rukopisných sporů nedovolilo, aby se od této formy kooperace matematiky s lingvistikou pokročilo k formě vyšší, totiž k účasti matematiky přímo na poznávacím jazykovědném procesu.

Historicky vývojová hodnota Seydlerova matematického příspěvku k rukopisnému sporu vynikne nejmarkatněji, zařadíme-li ji do časové řady oboru, do něhož svým charakterem patří a který dnes běžně označujeme jako matematická (kvantitativní) lingvistika. Nezjišťujeme však, zda ojedinělé práce předcházející Seydlerovi ve světovém kontextu mu byly známy. V r. 1847 upozornil ruský matematik V. JA. BUNJAKOVSKIJ (časopis *Sovremennik* 1847, tom III, oddíl II, 36–49) na možnost použití matematických metod v jazykovědě, v r. 1861 vyšla v Šanghaji práce W. GAMBLEHO, vypracovaná pro potřeby tiskařů (*Two Lists of Selected Characters Containing All in the Bible and Twenty Seven Other Books*). Užití matematických metod značně stimulovala platónská otázka; pro rozhodování o tom, které z dialogů a listů Platón skutečně psal, byla vypracována tzv. stylometrická metoda (CAMPBELL 1867, DITEENBERGER 1881 aj.) spočívající v zjiš-

tování frekvence některých formálních slov a hiátů. U nás na tyto práce později navázal FR. ČÁDA a FR. NOVOTNÝ (Listy filologické 1901 a 1906). Všechny práce, které jsou nám známy jako časoví předchůdci Seydlerovy studie, spokojovaly se však s konstatováním určitých jazykových jevů, tj. s deskriptivní statistikou, a na této úrovni byly i jejich interpretace. Na rozdíl od nich však Seydler provádí pravděpodobnostní hodnocení takových statistik vzhledem k různým možným hypotézám (používá Bayesova vzorce), zavádí si zmíněný koeficient koincidence a v tomto smyslu se odlišuje od svých předchůdců.

Ovšem z hlediska vlastního problému, tj. otázky pravosti RKZ, lze těžko považovat jeho postup za důkaz nepravosti; toho si byl nakonec sám Seydler dobře vědom. Jeho práci proto nelze zařadit mezi pokusy o určování autorství matematickými metodami, i když má s metodami atribuce textu několik společných kroků. Při atribuci textu je nutno obecně provést komplexní statisticko-stylistický rozbor anonymního (event. nepravého) textu i textů sloužících za porovnání, a to se zaměřením na nevědomé rysy autorova způsobu psaní.\*) Problém RKZ totiž není „klasické schéma atribuce textu, kdy máme k dispozici sporný text  $X$  a množinu  $M$  obsahující autory přicházející v úvahu jako možní autoři textu  $X$ . Složení množiny  $M$  vyplývá z dodatečných informací, většinou literárněhistorických. Rozhodujeme pak o autorství textu  $X$  vzhledem k množině  $M$ .

V případě RKZ jde nejprve o atribuci historické proveniencie textu a potom teprve eventuálně o přiřčení autorství. To znamená, že je nutno je porovnávat jak s jazykem a s jazykovými fenomény poetického tvaru pravých nejstarších památek středověké české literatury, tak s jazykem a s jazykovými fenomény poetického tvaru českého básnictví z počátku 19. století. Až potom – šlo-li by o atribuci textu určitému novověkému původci – uplatnilo by se navíc zmíněné „klasické“ schéma v patřičné modifikaci. Složitost kvantifikace takových poznatků – nutných k použití statistických metod atribuce textu – je zřejmá. Bylo by však nesprávné nevidět, že Seydler se o to svou podporou jazykovědných argumentů vlastně pokusil.

Proto důkaz o nepravosti RKZ musel přijít odjinud\*\*). Na Seydlerově práci je podle našeho názoru nutno ocenit především její myšlenku, tj. hodnotit jazykové jevy matematickými metodami. Právě v tom je jeho práce průkopnická a poučná i pro nás.

---

\*) Pro přehled takových postupů srov. P. VAŠÁK, *Statistika a sporné autorství*, Slovo a slovesnost 27, 1966, 364–370; též, *Metody ustanovenija spornogo autorstva (problema Barák - Neruda)*, Prague Studies in Mathematical Linguistics 3, 1972, 143–163.

\*\*\*) Srov. sb. redigovaný M. OTRUBOU, *Rukopis královédvorský a zelenohorský – dnešní stav poznání*, Praha 1970, který shrnuje a hodnotí výsledky, k nimž dospěly jednotlivé vědní obory zabývající se problematikou pravosti RKZ. Srov. též M. OTRUBA, *Mýtus a ritus. Pokus o sémantickou interpretaci obran pravosti RKZ*, Česká literatura 18, 1970, 213–277. Srov. M. IVANOV, *Záhada rukopisu královédvorského*, Praha 1970 (o novém chemickém zkoumání písma a pergamenu RK).