

# Historie matematické lingvistiky

---

## 2.1 Stručný historický nástin

In: Blanka Sedlačková (author): Historie matematické lingvistiky. (Czech). Brno: Akademické nakladatelství CERM v Brně, 2012. pp. 15–18.

Persistent URL: <http://dml.cz/dmlcz/402316>

### Terms of use:

© Blanka Sedlačková

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



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

## Kapitola 2

# Historie matematické lingvistiky

Počátky matematické lingvistiky klademe do 50. let 20. století, jedná se tedy o vědeckou disciplínu poměrně novou. Náznaky využívání matematických metod v lingvistice ale nacházíme již mnohem dříve, první stopy můžeme vysledovat dokonce už ve starověku. V této kapitole, která je těžištěm celé knihy, budou čtenářům představeny některé zajímavé „aplikace“ matematiky na jazyk z tohoto dlouhého prehistorického období (tedy z období před ustanovením matematické lingvistiky jako samostatné disciplíny). Vybrány budou takové aplikace, které v literatuře týkající se matematické lingvistiky nejsou uváděny vůbec nebo jsou zmiňovány jen velmi stručně bez bližšího objasnění. Dále bude čtenář seznámen s osobnostmi, metodami či pojmy, které sehrály důležitou roli při utváření matematické lingvistiky (a to jednak z období ještě před ustanovením matematické lingvistiky jako samostatné disciplíny či těsně po jejím vzniku). A konečně v závěru kapitoly se zmíníme o počinech, které jsou spjaty s českým kulturním prostředím a které v historii matematické lingvistiky zůstávají rovněž nezanedbatelné místo.

### 2.1 Stručný historický nástin

V této kapitole si shrneme běžně uváděné informace z historie matematické lingvistiky.

Vůbec poprvé nacházíme zmínku o využívání matematiky v jazyce ve starověku u starých Hindů, kteří z náboženských důvodů počítali slova v textu posvátné *Rgvédy*. Ve středověku byly velice oblíbené *obrazové básně*, které kladly důraz na mystiku čísla. Od 13. století se v Evropě začínají objevovat tzv. *spekulativní gramatiky*, které vedle popisu jazyka uplatňují i postupy filozofické, ale hlavně logické, založené na aristotelovské logice (Roger Bacon). Odtud je blízko k myšlence hledání tzv. *univerzálního jazyka* (Raimundus Lullus, John

Webster, John Wilkins, Francis Lodwick, Jan Amos Komenský aj.). Tito učenci často pracují s kombinatorickými pojmy a různými mechanickými pomůckami založenými na principu soustředných otáčivých kruhů. Ve většině případů jsou ale tyto středověké práce ovlivněny magií a astrologií, hlavní roli hraje *kabala*. Ta se rovněž zasloužila o vznik *steganografie* (tj. tajných písem), podoboru kryptografie, na počátku 16. století. Od 17. století pak začíná převládat čistě kombinatorický přístup k jazyku (John Wilkins, G. W. Leibniz). Racionalismus v jazykovědě se projevil sestavováním tzv. *filozofických gramatik*, platících obecně pro libovolný jazyk. Tyto snahy významně ovlivnily moderní lingvisty, např. Noama Chomského (více viz kap. 1.2), zakladatele *generativní a transformační mluvnice*. Podrobněji se aplikacím matematiky na jazyk v období starověku a středověku věnuje kapitola 2.2.

Zhruba od 19. století se v lingvistice začíná hojně uplatňovat využívání pojmu *frekvence* (neboli *četnost*). Zpočátku se jednalo o jednoduché počty různých jazykových jevů, které byly zpravidla vyvolány potřebami praxe (sestavování kazet s písmeny pro tiskaře, optimalizace klaviatury psacího stroje, Morseova abeceda, vytvoření optimální těsnopisné soustavy, zefektivnění jazykového vyučování, tvorba frekvenčních slovníků apod.). Podrobněji je pojem *frekvence* představen v kapitole 2.3.

Prvním, kdo zdůraznil využití matematických metod v lingvistice, byl s největší pravděpodobností ruský matematik Viktor Jakovlevič Bunjakovskij v roce 1847 (více viz kap. 2.4). V literatuře o matematické lingvistice bývá jako jeden z prvních vědců, který se zabýval kvantitativními jevy v jazyce, uváděn americký jazykovědec W. D. Whitney (1827–1894). V [6] na straně 126 čteme, že se „jako první lingvista vůbec zabýval frekvencí anglických hlásek a stal se tak předchůdcem dnešní statistické či kvantitativní lingvistiky.“ V historii kvantitativní lingvistiky ale můžeme nalézt práce mnohem starší, než jsou Whitneyho výzkumy z roku 1874. Zmíňme alespoň práce německých vědců Ernsta Förstemanna (1822–1906) a Augusta Schleichera (1821–1868). V kapitole 2.11 si představíme blíže dvě statistiky (hláskovou a lexikální), které vyšly dokonce již v roce 1831 v prvním českém vědeckém časopise *Krok*. O vážnějších pokusech zavést do lingvistiky kvantitativní hledisko můžeme hovořit až v období přelomu 19. a 20. století, což pravděpodobně souvisí se systematickým pronikáním matematiky i do jiných vědních oborů. Na konci 19. století mladogramatik Herman Paul chápe jazyk jako statistický průměr všech jazykových projevů jednotlivců a podobné kritérium zavádí i do fonetiky, kde invariantní hlásku považuje za statistický průměr jejích možných artikulací. Polský člen kazaňské školy Jan Baudouin de Courtenay (1845–1929) poukazuje na vhodnost využití v jazykovědě nejen metod elementární, ale i vyšší matematiky. Mezi osobnosti, které měly zásadní význam pro rozvoj matematické lingvistiky, patří v první polovině 20. století ruský matematik A. A. Markov (podrobně viz kap. 2.5). Ten aplikací statistických metod na text románu *Evžen Oněgin* od A. S. Puškina položil základy tzv. *markovských řetězců*. Další významnou osobností tohoto období je americký vědec G. K. Zipf, kterému je věnována kapitola 2.6 a který bývá považován za zakladatele kvantitativní lingvistiky. Jako první systematicky zkoumal a uveřejnil vztahy mezi frekvencí slov a jejich rankem

– tzv. *Zipfovy zákony*<sup>1</sup>. Později byly Zipfovy modely pojmově a matematicky vylepšeny a rozšířeny Benoîtem Mandelbrotem, světově uznávaným tvůrcem *fraktální geometrie*. U nás se aplikací Zipfových zákonů na češtinu zabývala Marie Těšitelová. Pro zajímavost zmiňme na závěr práci polského antropologa J. Czekanowského<sup>2</sup>, který matematickými metodami hledal důkazy o původním slovanském osídlení v oblasti Odry a Visly a jehož myšlenky se staly základem *taxonomické metody sestavování stemmatu* v textologii (blíže viz kap. 2.7).

Počátky matematické lingvistiky jako samostatné vědní disciplíny klademe do 50. let 20. století, zpravidla se vymezuje její vznik rokem 1948, kdy se konal *VI. mezinárodní lingvistický kongres* v Paříži. Zde byl totiž ustaven zvláštní výbor (*Committee on Quantitative Linguistics*), který se měl věnovat problémům kvantitativní lingvistiky. Jednatelem tohoto výboru byl zvolen český lingvista B. Trnka. Kromě toho bylo stanoveno množství témat (z lexika a gramatiky), která měla být zpracována analogicky pro různé jazyky (zejména pro potřeby typologického studia). Ačkoliv se kvantitativní lingvistika v řadě zemí v tomto období sice již rozvíjela, dělo se to bohužel spíše nekoordinovaně, a proto tato snaha zůstala ne zcela vyslyšena. Na vině je možná i to, že různé jazyky mají různé předpoklady pro využívání kvantitativních metod. Do tohoto období řadíme i vznik *teorie informace* (viz kap. 2.8), která se v nově vznikající kvantitativní lingvistice taktéž výrazně uplatnila. V 50. letech 20. století vznikla tzv. *glottochronologie* (neboli *lexikostatistika*), zajímavá aplikace matematiky v jazykovědě, která slouží k určování stáří jazyků. Podrobněji bude osvětlena v kapitole 2.9.

Podíváme-li se na uplatnění matematických metod při studiu jazyka v českých zemích, a zdůrazníme, že se jednalo zpočátku pouze o metody kvantitativní, můžeme s jistou mírou volnosti říci, že jako první pravděpodobně aplikoval tyto metody při studiu slovní zásoby již náš velký pedagog J. A. Komenský, kterému je věnována kap. 2.10. Ten ve svém díle *Janua linguarum reserata* ukázal, jak je možno znalosti o počtu slov v jazyce využít k efektivnímu rozvíjení slovní zásoby u žáků<sup>3</sup>. Mimo to se zabýval problematikou hledání *univerzálního jazyka* a seznámil se při svém pobytu v Anglii s naukou o tajných písmech čili *steganografií*. Velkou pozornost si zaslouží ojedinělý počín matematika a fyzika Augustina Seydlera, který roku 1886 aplikoval počet pravděpodobnosti na studium jazykových jevů za účelem potvrzení nepravosti tzv. *Rukopisů*<sup>4</sup> (více viz kap. 2.12). Zájem o využívání kvantitativních metod v lingvistice můžeme rovněž vysledovat v pražské lingvistické škole před druhou světovou válkou (základy této školy byly položeny v 30. letech 20. století). Za připomenutí

<sup>1</sup>Zejména viz *The Psycho-Biology of Language. An Introduction to Dynamic Philology*. Boston 1935; *Human Behavior and the Principle of Least Effort*. Cambridge 1949.

<sup>2</sup>Viz *Wstęp do historii Słowian*. Lvov 1927; nové vydání *Polska synteza slawistyczna w perspektywie ilościowej*. Warszawa 1947.

<sup>3</sup>Podrobněji viz [65].

<sup>4</sup>Viz [56], [57].

stojí práce V. Mathesia<sup>5</sup>, B. Trnky<sup>6</sup> (řešil rovněž i obecné otázky kvantitativní lingvistiky, včetně terminologických), dále Josefa Vachka<sup>7</sup> aj. I když to byli zaměřením především anglisté a pracovali tedy zejména s anglickým jazykem, výsledky své práce srovnávali s materiálem českým, popřípadě s materiálem z jiných jazyků. Jejich studium se týkalo hlavně oblasti fonologické a lexikální. Vedle lingvistů položili základy české matematické lingvistiky i pedagogové (podobně jako jinde ve světě). Tak například první frekvenční slovník češtiny [25] vznikl z podnětu pedagoga Václava Příhody a lingvisty bohemisty Vladimíra Šmilauera.

## 2.2 Starověk a středověk

První stopy užití kvantitativních metod v jazyce můžeme najít již u starých Hindů. Ti počítali z náboženských důvodů slova v textu posvátné *Rgvédy*, nejstarším z textů véd, který vznikl v letech 1500 až 1000 př. n. l. a byl sepsán někdy v letech 800 až 600 př. n. l. Jedná se o modlitební hymny, hymnické písně, které prostřednictvím staroindických světců (ršíů) vnukli lidem samotní bohové, aby je poučili o tom, jak je správně uctívat. Vědy zpočátku předávali bráhmani v ústní podobě z pokolení na pokolení. Aby při absenci jakéhokoliv písemného záznamu nedošlo k sebemenším změnám v textu Rgvéd, byl vytvořen důmyslný a velmi složitý systém určený k jejich uchování prostřednictvím ústního podání (védská metrika a časoměrný verš), který pracoval s přesnými počty slabik. Díky tomuto propracovanému systému se nám texty véd dochovaly v nezměněné podobě až do doby, kdy mohly být písemně zaznamenány.

Ve středověku se objevují aplikace na úrovni mystiky slov, čísel a obrazců. Můžeme sem zařadit středověké *obrazové básně* (*carmen figuratum*), ke kterým má blízko například Rabelaisův kaligraf *Božská láhev z Gargantuy a Pantagruela* (vydáno poprvé 1564). Odtud vede vztah přes Apollinairovy *kaligramy*, český *poetismus*<sup>8</sup>, *lettrismus* Isidora Isoua a *vizuální poezii* až do dneška (například tvorba počítačových obrazců vytvářených programátory). Ačkoliv tyto aplikace spadají do různých časových období, společná je jim právě již zmiňovaná mystika či hra.

Z předešlé kapitoly víme, že kvantitativní metody hrají významnou roli v *kabale* („*qabbalah*“ lze přeložit jako „tradice“), mohutném proudu hebrejského mysticismu, který považuje stvoření světa za jazykový jev. Nesprávně byla kabala v křesťanském světě spojována s černou magií, dokud nebyla přehodnocena humanisty. Kabala vychází z tradice výkladu *Tóry* (tj. *Pentateuchu*

<sup>5</sup> *O potenciálnosti jevů jazykových*. Věstník Královské české společnosti nauk, třída historická, 1911; nověji ve sborníku *U základů pražské jazykovědné školy* (vyd. J. Vachek). Praha 1970, s. 5–34.

<sup>6</sup> Viz [75].

<sup>7</sup> *Poznámky k fonologii českého lexika*. LF, 67, 1940, s. 395–402.

<sup>8</sup> Např. báseň *Počítadlo* a *Objevy* ze sbírky *Na vlnách TSF* od Jaroslava Seiferta, optická báseň *Adi* ve sbírce *Pantomima* od Vítězslava Nezvala aj.