

Historie matematické lingvistiky

2.3 Pojem „frekvence“

In: Blanka Sedlačková (author): Historie matematické lingvistiky. (Czech). Brno: Akademické nakladatelství CERM v Brně, 2012. pp. 28–48.

Persistent URL: <http://dml.cz/dmlcz/402318>

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>

2.3 Pojem „frekvence“

Přelom ve vztahu matematika versus lingvistika nastává s využíváním pojmu *frekvence* (*četnost*), jednoho z nejdůležitějších pojmů kvantitativní lingvistiky. Frekvence vyjadřuje počet určitého jevu v celku (viz např. tabulku 2.3). Jazykové jednotky ve všech rovinách (slova, slabiky, fonémy, hlásky, písmena, slovesné časy, pády, typy vět apod.) se vyskytují v různém počtu, mají různou četnost. Využívání znalosti frekvence určitých jazykových jevů bylo vyvoláno zejména potřebami praxe.

<i>Pořadí</i>	<i>Slovo</i>	<i>Frekvence</i>
1.	a	67 122
2.	být	43 148
3.	ten	37 280
4.	v(e)	33 679
5.	on	32 496
6.	na	27 753
7.	že	18 092
8.	s(e)	14 951
9.	z(e)	13 408
10.	který	11 692

Tabulka 2.3: Pořadí prvních 10 nejčastějších slov podle FSC¹²

Například tiskaři při sestavování kazet se zásobami písmen zjišťovali, že pro svou práci potřebují mnohem více takových písmen, která mají v daném jazyce frekvenci vysokou, než písmen s frekvencí nízkou (například v češtině by v takové kazetě musela být mnohem větší zásoba písmene „a“ než například písmene „ů“).

Rovněž stenografové při sestavování těsnopisných systémů přiřazovali z ekonomických důvodů nejjednodušší znaky těm jednotkám, které mají v daném jazyce nejvyšší frekvenci (podrobněji viz samostatná kapitola 2.3.2).

Frekvence hrála důležitou roli v pedagogické praxi – např. při hledání neefektivnějších metod vyučování cizím jazykům, zlepšování pravopisného systému apod. Již J. A. Komenský ve své učebnici *Janua linguarum reserata* (vydané poprvé roku 1631 v Lešně) zdůrazňuje znalost výskytu slov daného jazyka pro efektivnost vyučování, i když žádné přesné statistiky neuvádí.

¹²Běžně užívaná zkratka pro [25]. Tento slovník byl sestaven na základě 1 623 527 slovních výskytů ze 75 literárních děl a tiskovin různého charakteru z 8 různých slohově druhových skupin (krásná próza, poezie, literatura pro mládež, drama, odborná literatura, žurnalistika, vědecká literatura, mluvené projevy). Podrobněji popsáno v kapitole 2.3.1.

<i>Písmeno (znak)</i>	<i>Kód</i>	<i>Frekvence v angličtině (%)</i>	<i>Pomocná slova</i>
A	.-	8.167	akát
B	-. . .	1.492	blýskavice
C	-. .	2.782	cílovníci
D	-. .	4.253	dálava
E	.	12.702	erb
F	. . .	2.228	Filipíny; filiálka
G	-. .	2.015	Grónská zem; gnómóny
H	6.094	hrachovina
CH	----		chvátá k nám sám; chléb nám dává
I	. .	6.966	ibis; Ivan
J	. - -	0.153	jasmín bílý; junácká hůl
K	-. -	0.772	krákorá; království
L	. - .	4.025	lupíneček
M	- -	2.406	mává; mlází
N	- .	6.749	národ; nástup; Náchod
O	- - -	7.507	ó náš pán
P	. - .	1.929	papírníci
Q	- - -	0.095	kvílí orkán
R	-. .	5.987	rarášek
S	. . .	6.327	sekera; sobota
T	-	9.056	trám; tón; tuň
U	. . -	2.758	učení; uličník
V	. . -	0.978	vyučení; vyvolení; vinobraní
W	. - -	2.360	Waltrův vůz; wagón klád; dvojté vé
X	-. .	0.150	Xénokratés; Xénie má
Y	-. -	1.974	ý se krátí; ý se ztrácí; Ýkar mává
Z	-. .	0.074	zpíváme jen; známá žena; zrádná žena
1	. ----		
2	. . ----		
3	. . . -		
4 -		
5		
6	-		
7	- - . . .		
8	- - - . .		
9	- - - - .		
0	- - - - -		

Tabulka 2.4: Symboly pro kódování písmen a číslic v Morseově abecedě

Sestavování *Morseovy abecedy*, složené z teček a čárek, bylo rovněž založeno na pojmu frekvence. Toto kódování abecedy vymyslel americký vynálezce Samuel Morse a první telegrafické spojení uskutečnil roku 1844 mezi Washing-

tonem a Baltimorem. V roce 1918 byla původní abeceda dále zdokonalena. Konkrétně byla Morseova abeceda postavena na frekvenci písmen v angličtině. Písmenům nejčastěji se v angličtině vyskytujícím byla přidělena nejjednodušší sekvence teček a čárek. Proto například nejfrekventovanějšímu písmenu v angličtině „e“ odpovídá v Morseově abecedě znak „·-“ (tečka), písmenu „t“ odpovídá znak „-·“ (čárka) atd. Podrobněji viz tab. 2.4.

Podobně i na klaviatuře psacího stroje jsou nejčastější písmena na místech nejsnadněji dostupných nejobratnějšími prsty, tj. ukazováčkem a prostředníkem. I tato aplikace je ale optimálně sestavena pro angličtinu. Klaviatura českých psacích strojů vychází z této klaviatury anglické, je pouze doplněna o písmena v angličtině se nevyskytující a dále je provedena záměna písmen „z“ a „y“. Ale například polská matematická lingvistika v 60. letech 20. století provedla optimalizaci polských psacích strojů.

Frekvence písmen se využívala také při *dešifraci* textu. Vzhledem k tomu, že šifrováním a dešifrováním informací se zabývá disciplína nazývaná *teorie kódování* a není v našich silách popsat ani základní problematiku této disciplíny, omezíme se pouze na jeden zajímavý příklad z české literární historie. Básník Karel Hynek Mácha (1810–1836) si šifroval některé části svých intimních deníkových záznamů. Šifrování prováděl tak, že si sestavil vlastní abecedu, většinou odvozenou z řeckých písmen, a každé písmeno textu zaměnil symbolem z takto vytvořené abecedy. Situaci Mácha sice ještě poněkud ztížil tím, že češtinu kombinoval s německými slovy a větami a dále že každý druhý řádek (s jedinou výjimkou) psal pozpátku, ale při dešifraci pak v podstatě stačilo (zjednodušeně řečeno) sestavit tabulku výskytů Máchových symbolů a porovnat ji s obdobnou tabulkou sestavenou pro českou abecedu. Jednalo se vlastně o jednoduchou substituční šifru, kdy je jeden symbol nahrazen symbolem jiným. Řešení jednoduchých substitučních šifer bylo v arabském světě známo již v 9. století – Abú-Yūsuf Ya'qūb ibn Ishāq al-Kindī (801–873). Pro Evropu bylo řešení substitučních šifer objeveno mnohem později, a to někdy na začátku století čtrnáctého. Už tehdy se proto začala hledat vylepšení tohoto systému, která by umožnila zakrytí statistických závislostí v šifrovaném textu.

Všechny výše uvedené aplikace pojmu četnost se pohybovaly na úrovni pouhé deskripce, tj. něco se počítalo. Jako první na nedostatky pouhé deskripce a na vhodnost využívání vyšší matematiky upozornil roku 1847 V. Ja. Bunjakovskij (viz kap. 2.4). Nyní se ale u některých aplikací matematiky v lingvistice využívajících pojmu *frekvence* zastavme podrobněji. Nejprve si tu představíme problematiku frekvenčních a konkordančních slovníků a dále si ukážeme využití frekvence jazykových jednotek v těsnopise.

2.3.1 Frekvenční a konkordanční slovníky

S pojmem *frekvence* úzce souvisí *frekvenční* a *konkordanční* slovníky, které začaly hojně vznikat od konce 19. století. *Frekvenčním slovníkem* rozumíme víceúčelový seznam slov informující o jejich frekvenci, která bývá zjištěna na základě dostatečně rozsáhlého reprezentativně zvoleného materiálu. Zpravidla bývá v takovém slovníku uvedeno několik seznamů slov sestavených podle ur-

čitých hledisek, nejčastěji je to hledisko frekvenční (slova jsou seřazena podle klesající frekvence) a abecední (slova jsou uspořádána podle abecedy). Jednotlivá slova jsou zpravidla uváděna v základním tvaru (u substantiv, adjektiv, zájmen a číslovek to je 1. pád singuláru, u sloves infinitiv). Někdy bývají slova doplněna dalšími statistickými údaji. Pokud vytváříme pouze jednoúčelové seznamy slov na základě studia méně rozsáhlého materiálu (například dílo autora apod.), hovoříme o tzv. *frekvenčních seznamech*. Je-li ve frekvenčních slovnících navíc u slova uvedena informace, na jakém místě v excerpovaném textu se toto slovo nachází, mluvíme o *slovnících konkordančních*. Frekvenční seznamy, které bývají doplněny informací o tom, na jakém místě v textu se slovo nachází, bývají nazývány *indexy* (též *konkordance*).

První frekvenční slovníky vznikaly pro dobové praktické účely, například pro potřeby tiskařů¹³, konstrukci ekonomického těsnopisného systému¹⁴, zlepšení pravopisného systému¹⁵, tvorbu metod slepeckého čtení¹⁶, efektivnost vyučování cizím jazykům¹⁷ apod. Až poměrně pozdě se frekvenční slovníky začaly užívat jako důležitý nástroj k poznání systému a struktury slovní zásoby jednotlivých děl, autorů, funkčních stylů apod., dále pro účely strojového překladu, anotování a indexování textu z hlediska statistického. Zpočátku se frekvenční slovníky zpracovávaly ručně a uchovávaly pomocí excerpčních lístků. Aby byly frekvenční slovníky dostatečně vypovídající a odrážely skutečnou jazykovou situaci, je třeba zpracovat rozsáhlý materiál, což je ovšem ručně velmi pracné a hlavně zdoluhavé. Dnes tyto nedostatky umožňují úspěšně překonat počítače. V rámci počítačové lingvistiky se tak vytvořila samostatná disciplína, kterou nazýváme *korpusová lingvistika* a která se zabývá právě zpracováním rozsáhlých souborů jazykových dat (tzv. *korpusů*). V současné době existuje celá řada různých frekvenčních slovníků v různých jazycích, neboť tyto slovníky mohou být zajímavým zdrojem informací nejen pro lingvisty, ale i pro matematiky, informatiky, psychology, filozofy, pedagogy, metodiky, stenografy, kryptology apod.

Frekvenční slovníky se od sebe odlišují nejčastěji rozsahem zpracovávaného materiálu (ty novější mají rozsah od přibližně 500 000 slov až po několik milionů slov), výběrem excerpovaných textů (mluvené projevy, beletrie, poezie, dílo jednoho autora, díla určité literární skupiny apod.) či zpracováním (ruční, počítačové – dříve pomocí děrnoštitkových strojů). Vznikají různé speciální frekvenční slovníky (seznamy) – např. slovník děl (díla) určitého autora¹⁸, re-

¹³Gamble, W.: *Two lists of selected characters containing all in the Bible and twenty-seven other books*. Shanghai 1861.

¹⁴Käding, F. W.: *Häufigkeits Wörterbuch der Deutschen Sprache*. Steglitz 1897.

¹⁵Ayers, L. P.: *A Measuring Scale for Ability in Spelling*. NY 1915.

¹⁶Knowles, J.: *The London Point System of Reading for Blind*. London 1904.

¹⁷Eldridge, R. C.: *Six Thousand Common English Words: Their Comparative Frequency and What Can Be Done With Them*. Niagara Falls, NY 1911.

¹⁸*Slovar jazyka Puškina*. Moskva 1957; *Materialy k častotnomu slovarju jazyka Puškina*. Moskva 1963; *Slovník movy Ševčenka*. Kyjiv 1964; *Słownik języka Adama Mickiewicza*. Warszawa 1965; Sambor, J.: *Badanie statystyczne nad słownictwem (Na materiale Pana Tadeusza)*. Wrocław etc. 1969; *Goethe-Wörterbuch*. Berlin 1966; *Rečnik na ezika na Christo Botev I*. Sofija 1960; L. Benkó: *Juhász Gyula Költői nyelvének szótári feldolgozása*. Szeged 1957.

trográdní slovníky¹⁹, dialektologické slovníky²⁰, slovník literární skupiny apod.

Při zpracovávání frekvenčních slovníků se můžeme setkat s jistými problémy. Ty nastávají již při definování pojmu „slovo“. Nejjednodušší (a při počítačovém zpracování nejčastější) je chápání slova jako grafické jednotky, tzn. jako skupiny písmen vyskytující se mezi dvěma mezerami. Komplikace způsobují při tomto pojetí slova zejména složené slovesné tvary (*bude žehlit*), sdružená pojmenování (*vlčí mák, paví oko*), spřežky (často je lze psát dohromady i zvlášť – *vpodvečer* i *v podvečer*), stupňování přídavných jmen a příslovčí, číslovky (brát jako jednu lexikální jednotku celou číslovku či její části – *pět set*) či částice (dost dlouho se s tímto slovním druhem nepracovalo, zpravidla byly částice zařazovány k příslovcím, k přehodnocení dochází až v SSČ²¹), rozpoznání rozdílů mezi trpným rodem slovesným a spojením „být“ s přídavným jménem (*chlapec byl vychován* x *chlapec je nemocen*) apod. V těchto případech je třeba držet se zásady M. Těšitelové, a to té, že jednou přijatý způsob je třeba pokud možno všemi a všude dodržovat, aby bylo možno získané výsledky navzájem porovnávat.

Nyní si podrobněji představíme několik frekvenčních slovníků, které hrály významnou roli v historii kvantitativní lingvistiky. Nejpodrobněji se zastavíme u slovníků sestavených pro němčinu, neboť právě pro německý jazyk vznikl vůbec první frekvenční slovník na světě. Na tomto přehledu si ukážeme, jak se měnil obsah jednotlivých slovníků podle momentálních požadavků oboru. Orientačně se zmíníme o některých frekvenčních slovnících sestavených pro angličtinu, ruštinu, francouzštinu, španělštinu a rovněž pro slovenštinu a češtinu.

Němčina

Za vůbec první frekvenční slovník považujeme dvoudílný německý *Häufigkeitswörterbuch der deutschen Sprache* („Slovník četnosti výskytu německého jazyka“)²² stenografa F. W. Kädinga (1843–1928). Původně byl tento slovník vytvořen k těsnopisným účelům, podnět k jeho sestavení dal pracovní výbor německých těsnopisců v roce 1891. Výzkum probíhal více než 5 let na 110 pracovištích a byl zpracován materiál o rozsahu 20 milionů slabik (10 910 777 slov). Tento slovník se pak stal cennou pomůckou při kodifikaci těsnopisné soustavy Stolzeovy-Schreyovy v roce 1897. Byl založen na promyšlených obecných zásadách, z kterých později vycházely i další frekvenční slovníky. Slovník je složen ze dvou částí: v části první se uvádí údaje týkající se frekvence slov a slabik, v části druhé údaje o frekvenci písmen (ve skutečnosti se ale jedná o hlásky). V úvodu práce se správně předvídá, že tyto údaje mohou být důležité nejen pro těsnopisce, ale i pro lingvisty. Při tvorbě tohoto slovníku byl kladen důraz na „zákon velkých čísel“, tzn. na dostatečně velký rozsah materiálu, který by

¹⁹ Jsou to slovníky, v nichž jsou slova uspořádána v abecedním pořadí od posledního písmene k prvnímu. Využívány bývají například ve slovotvorbě - výskyt různých sufixů, prefixů apod.

²⁰ Jsou to slovníky zachycující slovní zásobu různých nářečí.

²¹ *Slovník spisovné češtiny pro školu a veřejnost*. Praha 1978. Další vydání: 2. opravené a doplněné vydání – 1994, 3. opravené vydání – 2003, 4. vydání – 2005).

²² Steglitz 1897, 643 s.

eliminovat počet slov s nízkou frekvencí, nadměrný výskyt slov závislých na tématu apod. Touto problematikou se později zabývala ve 20. století kvantitativní lingvistika (jednotka kvantitativní analýzy, výběr a rozsah materiálu apod.). Základní jednotkou je v Kädíngově slovníku slovo, přesněji tvar slova. Autor rozlišuje slova formální (řadí sem zájmena, adjektiva kvantitativní a číslovky, členy, pomocná a modální slovesa, nekvalitativní adverbia, předložky, spojky a citoslovce) a plnovýznamová (sem zařazuje substantiva, kvalitativní adjektiva a adverbia, konkrétní slovesa). Pro němčinu typické členy mají za následek nárůst počtu formálních slov, která jsou ale obecně nejčastější slova v různých jazycích.

Vedle svého prvenství je tento slovník významný i obrovským rozsahem ručně zpracovaného materiálu (10 910 777 slov – přesněji tvarů slov, z toho 258 173 různých tvarů slov). Dnes si zpracování tak rozsáhlého souboru bez pomoci počítačů vůbec nedovedeme představit. Pro potřeby stenografie byla slova dále členěna na slabiky (celkem asi 20 000 000 slabik) a na hlásky. K těsnopisným účelům bylo nutno slova rozkládat i z hlediska morfemického, tzn. na prefixy, kmeny, sufixy a koncovky. Značné úsilí bylo věnováno problematice složených slov – např. části složených slov a oddělitelné předpony byly zapisovány na různě barevné lístky. Všechny tyto slovtvorné prvky byly zaznamenány zvláště a potom řazeny podle frekvence, popřípadě podle abecedy. To znamená, že vedle frekvence slov jsou ve slovníku obsaženy i seznamy prefixů, sufixů a koncovek, výskyt samohlásek a souhlásek v prefixech, kmenech, sufixech a v koncovkách apod. Pozornost není věnována slovům o frekvenci 1 až 3, neboť pro těsnopisné účely nejsou podstatná. Je zde ale správně upozorněno, že pro některá jazyková studia mohou mít tato slova zvláštní význam.

Za zmínku stojí i výběr materiálu, neboť se jednalo o souvislé texty z těch oborů, v nichž se nejčastěji v tehdejší době užívalo těsnopisu. Celkem se jednalo o 16 tematických oblastí (právo, obchod, náboženství, medicína, historie, časopisy a knihy, vojenství, dopisy aj.). Vhodnost tohoto výběru se ukazuje, srovnáme-li tento slovník například s výběrem textů pro těsnopisné účely, který plánoval v letech 1940–43 První pražský spolek stenografů pod vedením B. Trnky, neboť tento český výběr se s tematikou textů v Kädíngově slovníku v podstatě shoduje, jen rozsah je značně menší.

Obecné zásady uplatněné při tvorbě Kädíngova slovníku se staly součástí řady pozdějších prací z kvantitativní lingvistiky (zejména v první polovině 20. století).

Na závěr si uvedme ještě několik zajímavých výsledků. Tak například prvních 15 nejfrekventovanějších výrazů představuje asi čtvrtinu celého slovního materiálu běžného německého textu, k pokrytí celé poloviny textu postačuje prvních 66 slov a prvních 320 výrazů pokrývá již téměř tři čtvrtiny textu. K běžnému dorozumění je ovšem potřeba znát podle výsledků různých vědců slov přibližně 850.

Kromě toho, že se jednalo o první frekvenční slovník vůbec, jednalo se rovněž o první dílo věnované lexikální a gramatické statistice. Ke Kädíngově slovníku se v pozdějších letech vracela celá řada lingvistů (např. Meier) a pedagogů (např. B. Q. Morgan). Ti však využili zpravidla jen část rozsáhlé slovní zásoby.

Zastavme se u některých pokračovatelů Kädinga podrobněji.

B. Q. Morgan²³ vypracoval v USA svůj frekvenční slovník němčiny pro potřebu jazykového vyučování, a to v době, kdy americká jazyková pedagogika opřela výuku mateřského jazyka, ale zejména jazyka cizího, o studium frekvence slov či jiných jazykových jednotek. Ve svém slovníku uvádí Morgan na rozdíl od Kädinga jen 2 400 nejčastějších německých slov. Protože podle něj Kädingův slovník nesl rysy svého původního určení, vyřadil ze seznamu uspořádaného podle klesající frekvence asi 320 slov, aby se dostal ke slovům, která jsou podle jeho mínění v němčině frekventovanější. Neuváděl rovněž další seznamy uvedené v Kädingově práci (frekvence hlásek, údaje o tvoření slov aj.). Lze pochopit, že Morgan pro potřeby jazykového vyučování využil jen některé ze seznamů, neakceptovatelné jsou ovšem subjektivní zásahy do původního seznamu.

Této chybě se chtěl vyvarovat H. Bakonyi²⁴, který sestavil rovněž pro pedagogické účely slovník nejužívanějších slov. Metodicky vycházel z Kädingova slovníku. Bakonyiův slovník obsahuje celkem 5 874 slov uspořádaných na základě frekvenčních údajů (absolutní a relativní frekvence) a k těmto slovům je připojen i tzv. *stupeň důležitosti*. Pro potřeby didaktické totiž nestačí někdy jen znalost slov nejčastějších, ale je třeba znát i slova nejnámější (tedy nejdůležitější). Známost slov se snažil zjistit na základě slovní zásoby dětského slovníku, kterým rozšířil materiál Kädingova slovníku. Na principu slovníku H. Bakonyiho u nás vznikly rovněž 2 slovníky, a to frekvenční slovník němčiny a angličtiny²⁵.

H. Meier²⁶ se pokusil o podrobné statistické roztřídění Kädingova materiálu na 12 stupňů častosti (*Häufigkeitsstufen*). Tyto stupně jsou dány frekvenčními intervaly, a to od frekvencí nejvyšších do frekvence nejnižší. Toto podrobné třídění ale vede k jisté nepřehlednosti celé práce.

Mezi další autory frekvenčních slovníků němčiny, kteří už ovšem nevycházeli ze slovníku Kädingova, patří H.-H. Wängler²⁷, který vydal pro němčinu frekvenční slovník hovorové horní němčiny. Slovník byl vytvořen na základě 80 266 slov ze severoněmecké oblasti zachycených v nenucené zábavě a pro srovnání byl též počet slov získán excerpcí různých textů z denního tisku a ze zábavných časopisů. Každé slovo pak bylo doplněno informací o frekvenci jednak v němčině hovorové, jednak v němčině psané.

Za vedení L. Hoffmanna vznikla v DDR na lipské univerzitě série frekvenčních slovníků odborných disciplín²⁸. Jedná se o speciální druh frekvenčního slovníku, který má obdobu v některých pracích sovětských²⁹. V podstatě jsou

²³Morgan, B. Q.: *German Frequency Word Book*. New York 1928.

²⁴*Die gebräuchlichsten Wörter der deutschen Sprache*. München 1934.

²⁵Kumprecht, K.: *Nejužívanější slova v němčině*. Praha 1936; Kumprecht, K. – Schulhofová, D.: *Nejužívanější slova v angličtině*. Praha 1947.

²⁶*Deutsche Sprachstatistik*. Hildesheim 1964.

²⁷*Rangwörterbuch hochdeutscher Umgangssprache*. Marburg 1963.

²⁸*Fachwortschatz Medizin. Häufigkeitswörterbuch. Russisch, Englisch, Francösisch*. Leipzig 1970; *Fachwortschatz Physik. Häufigkeitswörterbuch. Russisch, Englisch, Francösisch*. Leipzig 1970; *Fachwortschatz Chemie. Häufigkeitswörterbuch. Russisch, Englisch, Francösisch*. Leipzig 1973.

²⁹Piotrovskij, R. G. (red.): *Statistika teksta I*. Minsk 1969; Kiselevskij, A. I. (red.): *Statistika teksta II*. Minsk 1970.

to slovníky dvojjazyčné – například frekvenční slovník fyziky uvádí frekvenční seznam rusko-německý, anglicko-německý, francouzsko-německý. Seznamy obsahují kolem 1 100 slov a u jednotlivých lexikálních jednotek se uvádí jejich relativní frekvence. Slovníky mají sloužit jako metodická pomůcka pro žáky i učitele cizích jazyků. Za redakce L. Hoffmanna vyšel rovněž speciální sborník³⁰, v němž je obsažena problematika sestavování frekvenčních slovníků jazyka odborného stylu. Autor se v něm na základě zkušeností z práce na výše uvedených slovnících zabývá například metodologickými otázkami statistického zpracování vědeckých a technických odborných textů. Vedle toho jsou zde stati analyzující například slovník sovětských novin *Izvestija*, zvláštnosti ruských textů z oblasti marxisticko-leninské filozofie či z literární vědy.

I. Rosengrenová sestavila frekvenční slovník ze západoněmeckých novin³¹. Korpus byl získán reprezentativním výběrem z těchto novin v období jednoho roku (od 1. listopadu 1966 do 30. října 1967). Materiál z novin *Die Welt* má rozsah 6 545 600 slov, z novin *Süddeutsche Zeitung* 5 855 478 slov. Oba podkorpusy se člení na pět kategorií, v podstatě novinových rubrik: názory, politika, kultura, ekonomie a různé.

V roce 1972 vznikl první díl západoněmeckého frekvenčního slovníku německých odborných textů, jehož autorem je H. Erk³². Slovník přináší frekvenci sloves a jejich vazeb. Je vytvořen na základě materiálu o rozsahu 250 000 slov z 34 různých oborů. Ve slovníku se uvádějí pouze slova s výskytem nejméně 10. Jednotlivé lexémy jsou v seznamu doplněny *koeficientem distribuce a frekvence*. V roce 1975 vyšel druhý díl tohoto slovníku zaměřený na substantiva³³ a v roce 1982 díl třetí, který se zaměřuje na adjektiva, adverbia a ostatní slovní druhy³⁴.

Angličtina

K nejstarším anglickým frekvenčním slovníkům patří slovník L. P. Ayrese³⁵ z roku 1915, který vznikl na základě excerpcce 368 000 slov z obchodních a soukromých dopisů.

Pro účely vyučování angličtiny (a to jak mateřskému jazyku, tak jazyku cizímu) byl vytvořen jeden z neznámějších anglických frekvenčních slovníků, na němž se podílel E. L. Thorndike³⁶. Slovník tvoří 10 000 nejfrekventovanějších slov, která byla získána z materiálu o rozsahu 4 565 000 slov, pořízeného především z literatury pro mládež, z učebnic, rozmanitých příruček apod. Postupným rozšířením o dalších 5 milionů slov vznikla druhá verze slovníku³⁷

³⁰ *Fachsprachen und Sprachstatistik*. Berlin 1975.

³¹ *Ein Frequenzwörterbuch der deutschen Zeitungssprache Die Welt, Süddeutsche Zeitung*. Lund 1972.

³² *Zur Lexik wissenschaftlicher Fachtexte. Verben – Frequenz und Verwendungsweise*. München 1972.

³³ *Zur Lexik wissenschaftlicher Fachtexte. Substantive – Frequenz und Verwendungsweise*. München 1975.

³⁴ *Zur Lexik wissenschaftlicher Fachtexte. Adjektive, Adverbien und andere Wortarten. Frequenz und Verwendungsweise*. München 1982.

³⁵ *A Measuring Scale for Ability in Spelling*. New York 1915.

³⁶ *The Teachers Word Book*. New York 1921.

³⁷ *A Teachers Word Book of Twenty Thousands Words*. New York 1931–1932.

a ve spolupráci s I. Lorgem rozšířením o nových 4,5 milionů slov verze třetí³⁸. Podobně jako Kädingův německý frekvenční slovník i slovník Thorndikův byl zdrojem pro další práce, zejména z oblasti didaktické a lingvistické. L. Faucett a Itsu Maki vydali slovník opírající se o jazykový materiál 10 milionů slov³⁹, který vznikl spojením Thorndikova slovníku z roku 1921 a slovníku E. Horna⁴⁰, jenž byl sestaven ze soukromé a obchodní korespondence. Tento Hornův slovník se stal základem další práce, a to slovníku A. Hooda Robertse⁴¹, jehož cílem bylo zjištění frekvence v běžné americké angličtině.

Pozornost si zaslouží frekvenční slovník G. Deweye⁴² vytvořený na základě korpusu o rozsahu 100 000 slov zejména z publicistiky. Protože byl tento slovník určen k těsnopisným účelům, zjišťoval rovněž frekvenci slabik a hlásek.

Ojedinělým počinem bylo vytvoření prvního sémantického frekvenčního slovníku M. Westem⁴³ v roce 1953. Frekvence významů jednotlivých slov se uvádí v procentech, přičemž významy se rozlišují ve shodě se slovníkem *The Oxford English Dictionary*. Na Westův slovník navazovali rovněž další autoři. Například D. Wolf⁴⁴ vyšel z údajů ve slovníku *Shorter Oxford English Dictionary* (obsahuje 80 000 slov, která jsou opatřena údaji o počtu významů, druhu slov, o prvním dokladu apod.) a výsledky své analýzy srovnává právě se slovníkem Westovým (2 000 nejfrekventovanějších slov) a se slovníkem *Advanced Learners Dictionary* (28 000 slov). Sleduje zejména počet významů slov se zřetelem k prvnímu výskytu slova. Dokazuje tak celkem známý fakt, že čím je slovo starší, tím více má významů.

H. Kučera a W. N. Francis jsou autory frekvenčního slovníku současné americké angličtiny⁴⁵, který byl vytvořen na základě materiálu 1 014 232 slov z 500 textů (každý o rozsahu cca 2 000 slov), jež byly publikovány během jednoho roku (1963–1964). Při sestavování tohoto slovníku bylo již ke zpracování použito počítačů (stejně jako u slovníků Robertse a Wolfa).

Další anglický frekvenční slovník vznikl na území bývalého Sovětského svazu v letech 1971–1972. Jednalo se o třísvazkový frekvenční slovník slovních spojení v angličtině⁴⁶. Materiál byl získán ze 124 děl umělecké prózy (cca 7 milionů slov), 9 divadelních her (300 000 slov), 12 pramenů publicistických (300 000 slov) a 15 textů vědeckotechnických (500 000 slov). Východiskem při sestavování slovníku byl slovník Thorndikův-Lorgeovův. Slovník slovních spojení

³⁸ *The Teachers Word Book of 30 000 Words*. New York 1944.

³⁹ *A Study of English Word-Values Statistically Determined from Latest Extensive Word-Counts*. Tokyo 1932.

⁴⁰ *A Basic Writing Vocabulary*. Iowa City 1926.

⁴¹ *A Statistical Linguistic Analysis of American English*. The Hague 1965.

⁴² *Relativ Frequency of English Speech Sounds*. Cambridge 1923.

⁴³ *A General Service List of English Words with Semantic Frequencies and a Supplementary Word-List for the Writing of Popular Science and Technology*. London 1953.

⁴⁴ *Bedeutungshäufigkeit und ihr statistisches Verhalten*. Beiträge zur Linguistik und Informationsverarbeitung 1972, č. 22, s. 33–44.

⁴⁵ *Computational Analysis of Present-Day American English*. Providence, Rhode Island 1967.

⁴⁶ Volkova, N. O. – Ginzburg, R. Z. – Perebejnos, V. I. – Sankin, A. A. – Chidekel, S. S.: *Častotnyj slovar sočetajemosti sovremennogo anglijskogo jazyka*. 1. část, Moskva 1971; 2. část ve dvou svazcích, Moskva 1972.

v angličtině má význam nejen pro studium lexika, ale i gramatiky, stylistiky apod.

Ruština

Rovněž frekvenční slovníky ruštiny vznikaly jednak na území bývalého Sovětského svazu, jednak mimo něj.

Pro didaktické účely se zřetelem k vyučování ruštiny na českých základních školách zkoumala frekvenci slov v ruském jazyce A. Střížová⁴⁷. Excerpci sovětských učebnic autorka stanovila 2 060 slov označených jako nejčastější slova v ruštině. K témuž účelu zjišťoval frekvenci slov v ruštině F. Malír⁴⁸. Materiál získal excerpce sovětského tisku.

V USA byl ručně zpracován velký ruský frekvenční slovník H. H. Jossel-sona⁴⁹, který je založen asi na jednom milionu slov z psané ruštiny od roku 1830 do roku 1950. Podobně jako FSČ uvádí nejen frekvenci slov, ale i některých gramatických kategorií. Zřetel k frekvenci gramatických jevů, resp. jejich kategorií, je významným rysem frekvenčních slovníků vznikajících zhruba od 50. let 20. století.

Na půdě SSSR byl pořízen slovník E. A. Štejnfeldtové⁵⁰. Jedná se o seznam 2 500 nejfrekventovanějších slov, která byla zjištěna na základě materiálu 400 000 slov z 350 textů. Přestože byl určen pro potřeby školské praxe, má význam i pro lingvistiku. Byl přeložen do jiných jazyků, např. do němčiny⁵¹.

V tehdejší Leningradě vznikl frekvenční slovník současné ruštiny L. N. Zasorinové⁵². Materiál tvoří texty ze 4 stylistických skupin (po 25 %): umělecká próza, drama, populárně vědecké texty, publicistika. Vznik textů spadá do časového rozpětí 60 let.

Vedle toho vzniká v SSSR množství frekvenčních seznamů (či slovníků) pro speciální účely jednotlivých vysokoškolských učebních oborů, např. radioelektroniky, stavebních materiálů, fyzikální chemie, pro účely publicistické. Ukazují to například sborníky, které vyšly v Minsku za redakce R. G. Piotrovského⁵³ a A. I. Kiselevského⁵⁴.

Francouzština

První prací týkající se frekvence slov ve francouzštině byl *Le Vocabulaire d'un journal*, kterou pořídil Bony v letech 1920–1921 na základě jednoho čísla novin *Le Temps*.

⁴⁷ *Nejužívanější slova v ruštině*. Pedagogická revue 2, 1948, s. 335–359.

⁴⁸ *Rusko-český slovník nejdůležitějších slov pro četbu sovětského tisku*. Praha 1952.

⁴⁹ *The Russian Word Count and Frequency Analysis of Grammatical Categories of Standard Literary Russian*. Detroit 1953.

⁵⁰ *Častotnyj slovar sovremennogo ruskogo jazyka*. Tallin 1963.

⁵¹ Štejnfeldt, E.: *Häufigkeitwörterbuch der russischen Sprache*. Moskau 1965.

⁵² *Častotnyj slovar ruskogo jazyka*. Moskva 1977. Více informací o jeho koncepci nalezneme ve zprávě autorky (*Avtomatizacija i statistika v leksikografii*. Leningrad 1966.).

⁵³ Piotrovskij, R. G. (red.): *Statistika teksta I*. Minsk 1969.

⁵⁴ Kiselevskij, A. I. (red.): *Statistika teksta II*. Minsk 1970.

Pokus o sestavení frekvenčního slovníku francouzštiny učinil V. A. C. Henmon, který sestavil frekvenční seznam 4 000 slov na základě materiálu o rozsahu 400 000 slov. Tento seznam pak rozšířil a přepracoval G. E. Vander Beke⁵⁵. Vander Bekův slovník je založen na materiálu 1 147 748 slov a u jednotlivých slov uvádí jednak *frekvenční koeficient*, jednak tzv. *distribuční koeficient*, tj. v kolika textech se slovo vyskytuje. Tento slovník byl podobně jako Thorndikovy slovníky určen pro didaktické účely, zejména pro vyučování francouzštině jako cizímu jazyku. Na podkladu Vander Bekova slovníku vznikly práce další – u nás například pomůcka pro výuku francouzštiny od K. Kumprechta a A. Formánkové-Kubínové⁵⁶.

Frekvenční slovník současné francouzštiny vydal roku 1970 Al. Juilland jako třetí v sérii frekvenčních slovníků románských jazyků (vedle španělštiny a rumunštiny). Jeho spoluautory byli D. Brodin a C. Davidovitch⁵⁷. Korpus tvoří (podobně jako ve zmiňovaných frekvenčních slovnících španělštiny a rumunštiny) 500 000 slov z 5 žánrů (drama, beletrie, eseje, technická literatura a periodika). Slovník obsahuje u jednotlivých slov tyto koeficienty – *frekvenci*, *disperzi*⁵⁸ a *užití*⁵⁹. V obsáhlém úvodu jsou zmíněny nejen otázky týkající se problematiky této práce, ale i základní otázky lexikální statistiky vůbec (struktura korpusu, vymezení pojmu slova, koeficienty, interpretace výsledků aj.).

Zvláštní místo má frekvenční slovník projevů bývalého francouzského prezidenta Charlese de Gaullea⁶⁰, který je založen na 46 televizních politických projevech a obsahuje 6 195 slov (z toho asi 4 000 lexémů). Analyzují se dva typy televizních projevů, a to tzv. provolání a úvahové projevy.

Španělština

Vedle již zmíněného Juillandova slovníku vzniklo pro španělštinu několik dalších frekvenčních slovníků. Podle vzoru Vander Beka a Thorndika sestavil svůj frekvenční slovník M. A. Buchanan⁶¹. Opírá se o 1 200 000 slov ze 40 textů.

V 50. letech 20. století vyšly další dva frekvenční slovníky španělštiny. Slov-

⁵⁵ *French Word Book*. New York 1930.

⁵⁶ *Přehled nejčastěji užívaných slov ve franštině*. Praha 1935.

⁵⁷ *Frequency Dictionary of French Words*. The Hague 1970.

⁵⁸ *Koeficient disperze (D)*, též *rozptýlení*, zavedl A. Juilland a postihuje rozložení frekvence jednotky (slova) v jednotlivých textech korpusu, eventuálně v jejich částech. Lze jej vyjádřit vztahem

$$D = \frac{\delta}{m \cdot \sqrt{n-1}},$$

kde δ je směrodatná odchylka, m aritmetický průměr, n je suma všech frekvencí. Koeficient disperze se pohybuje v rozmezí 0–1. Při rovnoměrném rozdělení frekvence se blíží 0, při nerovnoměrném rozdělení se blíží 1.

⁵⁹ *Koeficient užití (U)* zavedl rovněž A. Juilland a lze jej vyjádřit vzorcem

$$U = \frac{F \cdot D}{100},$$

kde F je frekvence a D disperze.

⁶⁰ Cotteret, J. M. – Moreau, R.: *Le vocabulaire du général de Gaulle*. Paris 1969.

⁶¹ *A Graded Spanish Word Book*. Toronto 1927.

ník V. G. Hoze⁶², který obsahuje materiál 400 000 slov získaných ze soukromých dopisů, publicistiky a beletrie, je určen pro výuku španělštiny jako mateřskému jazyku. Stejný účel má i druhý slovník, jehož autorem je J. Rodriguez Bou⁶³. Tento slovník se opírá dokonce o materiál 7 066 637 slov.

Al. Juilland⁶⁴ vydal svůj frekvenční slovník španělštiny společně s matematikem E. Chang-Rodriguezem. Stavba slovníku je stejná jako u výše uvedeného frekvenčního slovníku francouzštiny, rovněž obsahuje u jednotlivých slov *koefficienty frekvence, užití a disperze*. Zpracován byl pomocí počítačů.

Slovenština

Rovněž první frekvenční slovník slovenštiny⁶⁵ (tradičně zkracováno na FSS) vznikl k těsnopiseckým účelům. J. Mistrík jej sestavil na základě materiálu 1 milionu slov, která byla získána excerpcí 60 celých textů 5 stylových skupin (drama, umělecká próza, poezie, žurnalistika, naučná literatura). Frekvenční seznam obsahuje celkem 9 568 lexikálních jednotek uspořádaných podle klesající frekvence, a to až do frekvence 3. Abecední seznam je tvořen 21 823 nejfrekventovanějšími slovy v slovenštině.

Čeština

Různé frekvenční statistiky se pro češtinu vytvářely již během 19. století. Jednalo se zpravidla o různé seznamy hlásek či písmen určené k různým praktickým účelům. Vůbec první jednoduchou hláskovou statistiku nacházíme roku 1831 v časopise *Krok*. Tato statistika si kladla za cíl vyvrátit tvrzení, že čeština je jazyk tvrdý s obtížnou výslovností, neboť je v něm obsaženo velké množství souhlásek. Podrobněji se o této statistice zmíníme v kap. 2.11. Statistické výzkumy lexika byly prováděny až později.

V roce 1911 byl založen Lexikografický archiv Ústavu pro jazyk český. V tomto archivu bylo shromážděno více než 9 miliónů excerpčních lístků. Dnes tyto lístky tvoří jeden abecední celek, ale původně těchto celků bylo asi 2 200. Každý tento celek byl samostatným dokumentem (dokumentem jazyka některého z našich předních spisovatelů, jistého časového úseku například v publicistice nebo odborné literatuře, tvorby určitého období apod.). Když se na základě tohoto materiálu začal zpracovávat slovník, bylo nutné všech těchto 2 200 abecedních celků sloučit v jeden, čímž byl ovšem ztracen materiálový podklad pro specializované práce. Na tomto příkladu se ukazují právě výhody strojního (dnes již počítačového) zpracování, neboť pomocí nich lze snadno pracovat pouze s částí jakéhokoliv celku (vyjmout například slovní zásobu autorů spřízněných lokálně, tematicky, generačně), tyto části dále zpracovávat (například uspořádat slovník podle stáří dokladů, což může být výhodné pro studium historie jazyka apod.), třídit podle různých kritérií (například podle

⁶² *Vocabulairo usual, comun y fundamental*. Madrid 1952.

⁶³ *Recuento de Vocabulario Español*. Puerto Rico 1952.

⁶⁴ *Frequency Dictionary of Spanish Words*. The Hague 1964.

⁶⁵ Mistrík, J.: *Frekvencia slov v slovenčine*. Bratislava 1969.

charakteru pramenů – próza, poezie, publicistika), provádět gramatická studia (třídění podle slovních druhů, gramatických kategorií atd.). Navíc všechny tyto seznamy by bylo možno velmi jednoduše doplnit o frekvenční údaje a vytvořit tak frekvenční slovník. Snadné by bylo rovněž sestavení např. slovníku retrográdního, v němž jsou slova abecedně uspořádána podle pořadí písmen od konce.

Od roku 1924 se provádějí první kvantitativní výzkumy v oblasti českého lexika, a to zejména pro účely didaktické či těsnopisné. Pro účely těsnopisné studoval na základě frekvence slov frekvenci hlásek ve slovenštině a v češtině A. Holas, dále například J. Sedláček⁶⁶. Z prací vzniklých k účelům didaktickým si zmínku zaslouží nepublikovaný *Slovník šestiletých dětí*, který z podnětu V. Příhody sestavili Th. Hamerský (na základě výzkumu ve Zlíně roku 1944) a M. Bartušková (na základě výzkumu v Praze v roce 1945).

Roku 1938 vznikla první kvantitativní práce z oblasti jazyka publicistiky⁶⁷, a to na základě excerpce jednoho čísla deníku *Národní osvobození* (celkem 42 448 slov). Marie Těšitelová⁶⁸ roku 1948 provedla kvantitativní rozbor knihy Karla Čapka *Život a dílo skladatele Foltýna*.

První frekvenční slovník češtiny byl vydán roku 1961 v Praze pod názvem *Frekvence slov, slovních druhů a tvarů v českém jazyce* (standardně se označuje zkratkou FSC). Autory byli J. Jelínek, J. V. Bečka a M. Těšitelová. Původně byl slovník určen pro účely didaktické, ale toto své určení daleko překročil, takže slouží i ke studiu kvantitativních poměrů v češtině vůbec. Slovník vznikl ručně po dobu více než 20 let na základě 1 623 527 slov získaných ze 75 textů. Celkem je v něm obsaženo 54 486 slov různých. Kdyby autoři neměli k dispozici četné pomocníky, dílo by vůbec nemohlo vzniknout. FSC rozlišoval osm funkčních stylů – čtyři umělecké (beletrie, básně, literatura pro mládež, dramata), čtyři odborné (populárně vědecká literatura, noviny, vědecká literatura, mluvené projevy). Zastoupení bylo ovšem nerovnoměrné: beletrie zaujímala 30 % dokladů, mluvené projevy jen 6 %. Vedle frekvence slov se ve FSC uvádí například i frekvence některých gramatických kategorií (např. slovních druhů, pádů substantiv apod.).

Ve stejném roce vznikl spoluprací Oddělení pro teorii strojového překladu Filosofické fakulty Karlovy univerzity a strojně početní stanice Vysoké školy ekonomické v Praze frekvenční slovník matematických textů. Slovník obsahoval 51 341 slov představujících asi 3 050 různých lexikálních jednotek. Byl zpracován pomocí děrných štítků. Pro účely strojového překladu byl na stejných pracovištích vytvořen frekvenční slovník anglických elektrotechnických textů. Rovněž byl sestaven pomocí děrných štítků. Na štítek se vedle anglického slova uváděl i slovní druh a český ekvivalent (překlad), třídít slova pak bylo možno jak podle slov anglických, tak podle českých, což je jistě významná pomůcka při řešení homonymie a synonymie.

⁶⁶ *Poměry slov a jejich částí v češtině, získané rozbořem 18 000 slov*. Těsnopisné rozhledy 2, 1924, s. 65–66.

⁶⁷ Korejs, J. – Korejsová, H.: *Slovník našich novin*. Praha 1938.

⁶⁸ *Frekvence slov a tvarů ve spise „Život a dílo skladatele Foltýna“ od Karla Čapka*. Naše řeč 32, 1948, s. 126–130.

Z dalších prací podobného charakteru jmenujme alespoň v roce 1969 v Praze vydaný *Konkordanční a frekvenční index k Slezským písním Petra Bezruče*, i ve světovém kontextu ojedinělé dílo *Biblická konkordance*⁶⁹ a další práci J. V. Bečky, který zkoumal složení českých odborných textů technického zaměření na materiálu 175 000 slov ze 7 vědních oborů: lékařství, technická chemie, zemědělství, strojírenství, báňský a hutní průmysl, elektrotechnika, stavitelství⁷⁰. Vedle úvodu, v němž se probírají základní otázky techniky práce i některé výsledky, obsahuje práce frekvenční seznam slov a slovních skupin.

2.3.2 Těsnopis

Jednou z oblastí, která z praktických důvodů využívala výzkumy frekvence jazykových jednotek, byl již od 19. století *těsnopis*.

Těsnopis (též *stenografie*, *rychlopis*, *tachygrafie*) je písmo umožňující psát rychlostí mluvené řeči. Podle B. Trnky se těsnopis liší od obvyklé grafické reprodukce řeči časovou a prostorovou úsporností, „*má tudíž k obyčejnému písmu poměr písma pomocného, prostředkujícího mezi ním a mluveným, po případě myšleným slovem.*“ ([76], str. 7)

Novodobé rukopisné těsnopisné soustavy se většinou rozdělují na dvě skupiny:

- 1) *soustavy geometrické*, ve kterých jako znaky vystupují většinou geometricky nejjednodušší tvary (přímka v různých polohách, kružnice a její části, smyčka, klička, tečka apod.) a tvary vzniklé jejich spojováním; tyto soustavy začaly vznikat na přelomu 16. a 17. století v Anglii.
- 2) *soustavy kurzívní*, které se snaží co nejvíce přiblížit obyčejnému psacímu písmu pravidelným střídáním pohybů shora dolů a zdola nahoru, používají i znaků geometricky složitějších. Tyto systémy začaly vznikat na počátku 19. století zejména v Německu. Důraz byl kladen na to, aby tyto systémy byly co nejplynulejší.

Vedle rukopisných těsnopisných systémů vznikly v moderní době i soustavy strojové (klávesnicové), které jsou rozšířeny zejména v USA a západní Evropě.

Za předstupeň těsnopisu můžeme považovat tzv. *akropolský kámen* z doby kolem 350 př. n. l. (objeven roku 1883), na němž byly nalezeny zásady pro tvorbu zkrácených znaků. Kolem roku 300 př. n. l. vznikly tzv. *delfské desky*, na kterých neznámý autor uvádí seznam znaků nahrazujících skupiny dvou až tří hlásek.

Za objevitele těsnopisu je považován Marcus Tullius Tiro (103–4 př. n. l.), nejdříve otrok, později Ciceronův tajemník, který k označení jednotlivých slov užíval tzv. *not* (odtud název „*tironské noty*“). Tiro vyšel ze starší soustavy

⁶⁹ Bič, M. – Souček, J. B.: *Biblická konkordance*. Ústřední církevní nakladatelství Praha, 1954–1967; též Bič, M. – Souček, J. B.: *Biblická konkordance (3 sv.)*. Kalich, Praha 1961, 838 s.; 1963, 777 s.; 1967, 820 s.

⁷⁰ Bečka, J. V.: *Lexikální složení českých odborných textů technického zaměření I*. Praha 1973; Týž: *Lexikální složení českých odborných textů technického zaměření II*. Praha 1975.

značek, kterými se zkracovala slova v obyčejném písmu. Poprvé byl tento těsnopisný systém použit v Římě 5. prosince 63 př. n. l. k zapsání Catonovy obžalovací řeči proti Catilinovi. Těsnopis se užíval i ve vojsku ve funkci tajného písma. Ve starověkém Římě se zpracovávaly stenografické slovníky, učebnice a prováděla se různá zlepšení. Po rozpadu Římské říše roku 476 n. l. se těsnopisný systém přenesl i do Galské a Francké říše. Užíval se zejména v kláštorech, ale jeho složitost (velké množství not) způsobila, že se v průběhu 10. a 11. století užívat přestal. V Itálii se vytvořil slabikový těsnopis (6. – 8. století, novější verze 8. – 11. století). Ačkoliv byl kratší než obyčejné písmo, byl méně rychlý než slovní písmo, což bylo důvodem k jeho zániku.

Ve středověku se pak výrazně uplatnila *zkrácená latina*. Za autora tohoto zkráceného písma je považován mnich John of Tilbury, který v 11. století na podnět arcibiskupa z Canterbury Thomase Becketa zkoumal systém *tíronských not*. Ačkoliv vyzdvihl systém tohoto zkráceného písma, konstatoval obtížnost těchto not. Sestavil tedy nové latinské slovní písmo, v němž kmeny označoval hlavními notami, koncovky notami pomocnými.

Na vývoj novodobého těsnopisu mělo velký vliv tajné písmo (*steganografie*). Autorem jednoho z nejpropracovanějších systémů byl londýnský lékař, později venkovský farář Timothy Bright (1551–1615), který roku 1588 vydal *Characterie* (tj. umění krátce, rychle a tajně psát pomocí charakterií), první knihu anglického rychlopisu. Podobnou soustavu o dva roky později vypracoval Peter Bales (1547–1610?) v knize *The Art of Brachygraphy*. Brzy je ale zastínil John Willis (1575–1625). Tento londýnský farář vydal roku 1602 knížku o 100 stranách malého formátu *Art of Stenographie* (jen do roku 1647 byla vydána celkem 14krát). Autor zde také poprvé použil označení „*těsnopis*“⁷¹. Zavedl pevnou abecedu, jejíž znaky byly vytvořeny z přímků a oblouků v různé poloze, tečky, kruhu a smyčky (první geometrická těsnopisná soustava). Při volbě znaků přihlížel údajně i k frekvenci.

Při své návštěvě Anglie se s Willisovou těsnopisnou soustavou seznámil i Jan Amos Komenský. V jeho latinsky psaném dopise *Ad amicos Lesznae in Polonia agentes* z Londýna 8. října 1641 pak nacházíme nejstarší zmínku o těsnopise u českého autora (viz [31]): „*Valná část mladších i starších mužů zaznamenává si kázání perem, a to doslova. Bylo tu totiž před 30 lety (za Jakuba) vynalezeno a již i na venkov proniklo rychlopisné umění, jež oni nazývají steganografie, jímž (ne pomoci písmen, nýbrž značek značících celá slova) rychlost jazyka rukou napodobují. Učí se pak onomu umění ve městech skoro všichni, jakmile se naučili obyčejnému písmu ve škole, přidávajíce asi rok k naučení steganografie.*“ Komenský se tu patrně seznámil s Brightovou, popř. Balesovou soustavou tajného písma, neboť v textu užívá pojem steganografie (tajnopis), což by zdůvodňovalo i existenci zvláštních značek pro celá slova. Ve svém spise *De rerum humanarum emendatione consultatio catholica* pak Komenský uvádí nejen návrh na jednotný mezinárodní dorozumivací jazyk (tzv. *univerzální jazyk*), ale je rovněž autorem skici těsnopisné soustavy pro tento jazyk.

Od počátku 19. století se začíná významně uplatňovat v těsnopisu pojem *frekvence*. Těsnopisci si začali uvědomovat důležitost jazykového popisu a právě

⁷¹Viz [48].

těsnopis byl podnětem pro různé jazykové rozbory. Například zakladatel německého kurzívního těsnopisu Franz Xaver Gabelsberger (1789–1849) získával při sestavování své těsnopisné soustavy podněty jednak studiem těsnopisné literatury, jednak praktickou stenografickou činností. Při volbě znaků pak přihlížel především k frekvenci – přiděloval častým hláskám znaky jednodušší a psatelnější⁷². Těsnopis byl inspirací pro první frekvenční slovník na světě, a to pro frekvenční slovník F. W. Kädinga z roku 1897⁷³. Statistická šetření publikovaná v tomto slovníku a týkající se frekvence slov, slabik i hlásek v německé řeči se stala cennou pomůckou při kodifikaci Stolzeovy-Schreyovy soustavy v roce 1897.

Kvantitativní vlastnosti jazyka za účelem sestavení optimální těsnopisné soustavy zkoumal rovněž Francouz Jean-Baptiste Estoup (1868–1950). Ve Francii založil *Institut International de Métagraphie Duployé* (odnož těsnopisné školy Duployé, jedné z nejrozšířenějších těsnopisných soustav ve Francii). Při vyučování těsnopisu se uplatňují jeho cvičení, která obsahují kombinace vysoce frekventovaných slov (tzv. *Estoupovy stenografické gamy*). S jeho jménem je spjat rovněž tzv. *Zipfův-Estoupův zákon* (více viz kap. 2.6).

Soustava GESS (= Gosudarstvenaja jedinaja sistema stenografii) N. N. Sokolova, která byla 10. června 1933 sovětskou vládou zavedena jako jediná státní těsnopisná soustava, se opírala rovněž o rozsáhlý vědecký výzkum v oblasti biomechaniky, jazykovědy, matematiky apod.⁷⁴

Na Slovensku přední místo zaujímají práce Jozefa Mistríka. Ten publikoval celou řadu stenografických prací a jeho dílo je dalším příkladem toho, jak se původně stenografický výzkum⁷⁵ stal základem práce s daleko širším záběrem, a to frekvenčního slovníku slovenštiny⁷⁶. Spoluprací J. Mistríka (později Ladislava Bojnanského) s českými těsnopisci K. Matouškem a M. Matulou bylo do slovenského těsnopisu z češtiny přeneseno tzv. krácení II. stupně.

Na základě požadavků stenografů připravili velký frekvenční slovník polští jazykovědci a rovněž například v Bulharsku vznikla řada prací z oblasti statistiky písmen vyvolaná konkurzem na novou bulharskou těsnopisnou soustavu (viz [31]).

Pomineme-li zmínku o steganografii z pera Jana Amose Komenského z roku 1641, první pojednání o těsnopise na našem území roku 1834 publikuje Jan Dalibor Kopecký v pražském časopise *Květy*. Jedná se o stručné seznámení s těsnopisem vůbec a zvláště s Gabelsbergerovou soustavou používanou v Německu, ve kterém vyslovuje přání, aby toto umění bylo přeneseno a pěstováno i u nás.

⁷²Zda tak činil na základě nějakého konkrétního statistického výzkumu, se v literatuře neuvádí.

⁷³*Häufigkeitswörterbuch der deutschen Sprache*. Steglitz 1897.

⁷⁴Zásady vyložil Sokolov v díle *Těoretičeskije osnovy gosudarstvennoj jedinoj systemy stenografii*. Moskva 1949. Tato kniha patří k jedněm z nejpracovitějších děl věnovaných základním principům těsnopisu.

⁷⁵*Slovenský stenograf VIII*. 1954, s. 85–86 a 105–107; *Stenografia I*. Bratislava 1955, s. 195–223.

⁷⁶J. Mistrík: *Frekvence slov v slovenčine*. Bratislava 1969.

První českou těsnopisnou soustavu vypracoval Hynek Jakub Heger (1808–1854) ve 40. letech 19. století jako převod německé soustavy F. X. Gabelsbergera, a to nejen pro češtinu, ale i pro hlavní slovanské jazyky (tzv. *těsnopisný panslavismus*). Ačkoliv měl Heger nemalý podíl na rozšíření těsnopisu do Čech, neměla jeho soustava dlouhého trvání. Heger do původní soustavy Gabelsbergerovy z úcty před autorem vnašel totiž jen málo změn a výběr znaků ne úplně vyhovoval češtině.

Roku 1859 vznikl pod vedením Jindřicha Fügnera *První pražský spolek stenografů*, který roku 1861 vypsal soutěž na nejlepší převod Gabelsbergerovy soustavy do češtiny (vítěz měl získat 25 dukátů ve zlatě). Z pěti návrhů ale porota žádný nedoporučila. Práce na převodu přesto pokračovaly a na Nový rok 1863 vyšla malá publikace *Těsnopis český* jako výsledek práce sedmi členů spolku stenografů (Eduard Novotný, František Brzobohatý, Josef Faigl, A. Konrad, Josef Novotný, Alois Stáně a František Tauer). Autoři přihlédli k některým zvláštnostem českého jazyka, i když ne úplně důkladně. Chyběly podklady o frekvenci hlásek a jejich kombinací. Ačkoliv se tato práce setkala nejprve s kritikou německých odborných časopisů, byla později znovu vydávána a přepracovávána, a to až do roku 1921, kdy vyšlo 14. vydání této práce.

Autorem první původní soustavy českého těsnopisu byl kněz František Gába (1829–1896), který tvrdil, že český těsnopisný systém by měl být založen na osobitých rysech slovanských jazyků (zejména češtiny) a nemůže být jen převodem německé soustavy. Zdůrazňoval tyto dva požadavky:

- 1) Písmena by měla být v rámci možností obrazem zvuku, tzn. artikulačně (foneticky) podobné hlásky by měly být zachyceny graficky podobnými znaky. Tento princip se uplatnil později v Trnkově těsnopisu, který byl spjat s vývojem funkčního hláskosloví ve 20. a 30. letech 20. století.
- 2) Tvorbě soustavy by měl předcházet rozsáhlý jazykový rozbor, zaměřený zejména na poznání kvantitativních vlastností jazyka (zvláště slovníku a grafematické roviny). Počátkem 50. let 19. stol. sám Gába údajně prováděl kvantitativní výzkum, na jehož základě se pokoušel sestavit takovou soustavu, ve které by graficky nejjednodušší znaky odpovídaly nejčastěji se vyskytujícím písmenům.

I když se v praxi nadále osvědčoval převod soustavy Gabelsbergerovy, Gábova myšlenka na původní českou těsnopisnou soustavu nezanikla. Roku 1864 český lékař Jan Bareš vydal spisek *Základy slovanského rychlopisu*, ve kterém se pokoušel vytvořit takovou těsnopisnou soustavu, kterou by bylo možno převést do ostatních slovanských jazyků. „*Konstrukci slovanské abecedy předcházela u Bareše výpočet frekvenčního poměru hlásek v češtině, ovšem na velmi omezeném materiálu.*“ ([48], str. 91) V soustavě se ale vyskytovala řada nedostatků způsobených autorovými neznalostmi slovanských jazyků i nedostatkem praktických zkušeností, byla proto komisí pražského spolku zamítnuta.

Antonín Krondl ve svém referátu *Kterak vyhovuje převod soustavy Gabelsbergerovy hláskoslovné povaze jazyka českého* na třetím sjezdu českých stenografů (konaném 20. a 21. května 1888 v Brně) ukázal, že při sestavování

původní české těsnopisné soustavy bude poznání kvantitativních charakteristik jazyka nezbytné a bude třeba opírat se o přesně zjištěnou *iteraci* (= opakování) a zvláště kombinaci českých hlásek. Sám provedl řadu frekvenčních výzkumů pro práci na své původní české těsnopisné soustavě.

Opomenutá zůstala rukopisná práce Josefa Sedláčka z let 1913–1917 *Základní studie k českému těsnopisu*, která obsahuje jednak frekvenční seznam nejčastějších slov sestavený ze souboru 100 000 slovních výskytů, jednak frekvenci hlásek, hláskových skupin a slabik ze souboru 18 000 slov⁷⁷. Přesněji se jedná o kvantitativní charakteristiku psaných textů stylu veřejných projevů ze začátku 20. století⁷⁸.

Statistické výzkumy prováděl i František Hradilík (1871–1939), lektor těsnopisu na brněnské technice. Zúčastnil se veřejné soutěže vyhlášené ministerstvem školství dne 8. července 1920 na původní soustavu československého těsnopisu a jako přípravné práce k ní konal frekvenční výzkumy hlásek a hláskových skupin na 10 000 českých slovech. Iterace a kombinace byly rovněž vědeckým základem soustavy inženýra Svojmíra Mikulíka a PhDr. Aloise Herouta, která se stala vítězem této soutěže na původní český těsnopis (15. června 1921 se porota jednomyslně usnesla na jejím vítězství z celkem 14 soustav). Tato soustava (tzv. *soustava Heroutova-Mikulíkova*) byla jmenována oficiálním českým těsnopisem. 23. června 1921 ministr školství schválil zavádění nové soustavy do škol, přičemž zkoušet podle ní se mělo při učitelských státních zkouškách od letního termínu zkušební roku 1922. K jejímu rozšiřování a zdokonalování byla zřízena 6. prosince 1921 státní instituce *Státní ústav těsnopisný*.

Soutěže na původní českou těsnopisnou soustavu se zúčastnil mimo jiné i profesor anglické filologie Karlovy univerzity v Praze PhDr. Bohumil Trnka, jeden z nejvýznačnějších světových průkopníků v oblasti funkční fonologie. Trnka se pokusil o vybudování důsledné a lingvisticky odůvodněné těsnopisné soustavy. Byla založena na soustavě distinktivních rysů fonologického systému a jednotlivým hláskám byly přiřazeny grafické znaky podle hláskové iterace (výskytu) a kombinace. To znamená, že se Trnka opíral nejen o kvalitativní výzkum jazyka (reprezentovaný strukturalismem, zejména pak N. S. Trubeckým, V. Mathesiem, R. Jakobsonem, B. Havránkem a jinými členy Pražského lingvistického kroužku), ale i o výzkum stránky kvantitativní. S prvním návrhem⁷⁹ své soustavy vystoupil již v roce 1914, přepracovanou ji předložil do ministerské soutěže. Bohužel Trnkova těsnopisná soustava nenašla většího ohlasu, neboť se její návrh objevil až po převládnutí systému Heroutova-Mikulíkova, který je s některými změnami oficiálním těsnopisným systémem dodnes. Vý-

⁷⁷Rozbor práce je zahrnut v knize M. Matuly: *Teorie a praxe těsnopisu I*. Praha 1958, s. 116n.; též poznámka u B. Trnky: *Pokus o vědeckou teorii a praktickou reformu těsnopisu*. Praha 1937, s. 33n.

⁷⁸Pro srovnání s dnešními poměry viz M. Ludvíková – J. Kraus: *Kvantitativní vlastnosti soustavy českých fonémů*. SaS 27, 1966, s. 334n. Sedláček se od nich mírně odlišuje vymezením sledovaných jednotek, např. nerespektováním ztráty znělosti na konci slova, rozlišováním *i/y* apod.

⁷⁹Podrobně svou koncepci vyložil v [76]. V předmluvě Trnka vyzdvihuje těsnopis jako důležitou dovednost a dále oceňuje jeho vliv na vývoj fonetických teorií či jeho důležitou roli v literární historii.

znam Trnkova návrhu původní těsnopisné soustavy ale spatřujeme v propojení strukturní fonologie, těsnopisu a kvantitativní lingvistiky.

Je tedy vidět, že pro potřeby stenografie byla provedena řada různých jazykových statistik. Přestože tyto statistiky zpracovávaly jen omezený materiál a neodpovídaly dnešním požadavkům na jazykový rozbor, jsou spolu s problematikou sporného autorství považovány za nejstarší zdroje matematické, přesněji kvantitativní lingvistiky.

Pro novější český těsnopisný výzkum je typická spolupráce stenografů s vědci z příbuzných oborů – jazykověda, matematická lingvistika, fyziologie, psychologie psaní apod. Zmíňme zde alespoň Miloše Matulu⁸⁰. Tento vědecky aktivně pracující matematik se zájmem o otázky jazykové je rovněž vynikající praktik v oblasti těsnopisu – je držitelem českého rekordu, kdy v roce 1957 dosáhl rychlosti 200 slov za minutu v pětiminutovém diktátu s odstupňovanou rychlostí. Jen pro představu uvedme, že pro zápis běžné řeči stačí rychlost zápisu 70 slov za minutu, pro rychlejší přednes či například televizní zprávy je to 80 až 100 slov za minutu. Píšeme-li běžným písmem, bývá rychlost zápisu okolo 20 až 40 slov za minutu. Největší Matulův význam ale spatřujeme v rozpracování *teorie krácení slov* (oproti dosavadnímu hledání nejnvýhodnější abecedy založené na studiu frekvence). Společně s K. Matouškem vytvořil tzv. *silné krácení* (také *krácení II. stupně*), založené na slovotvorných vlastnostech jazyka, které stenografům umožňuje psát vyšší rychlostí. V první fázi prací na krácení II. stupně posloužila autorům neprávem opomenutá práce Josefa Sedláčka *Základní studie k českému těsnopisu*, později se jako potřebné ukázaly frekvenční statistiky slov a slovních skupin v projevech hospodářských a politických⁸¹. Na závěr ještě připomeňme, že vedle *krácení II. stupně* se M. Matula zabýval analýzou a měřením procesu stenografování, při němž využíval statistického průzkumu a matematických metod. Z českých lingvistů, kteří se věnovali problematice těsnopisu, jmenujme alespoň J. Krause.

Různé frekvenční výzkumy se prováděly i v pozdějších letech, zejména pod záštitou Státního ústavu těsnopisného, který se v teoretické oblasti zaměřoval především na usnadnění naučitelnosti těsnopisné soustavy Heroutovy-Mikulíkovy. Výsledkem této snahy byla reforma Heroutovy-Mikulíkovy soustavy v roce 1961, která byla založena na několik let trvajících výzkumech frekvenčních, srovnávacích a grafických. Ústav měl rovněž svůj podíl při položení základů *krácení II. stupně*.

V souvislosti s kritickým hodnocením Heroutovy-Mikulíkovy těsnopisné soustavy a s novou vlnou zájmu o otázky frekvence slov, písmen a jejich kombinací prováděl při práci na své nové těsnopisné soustavě STENOP rozsáhlé statistiky například Josef Horák (vedle podrobného studia fyziologie a psychologie psaní)⁸². Ačkoliv v druhé polovině 20. století vznikly nové návrhy na modernější

⁸⁰Viz Matula, M.: *Teorie a praxe těsnopisu*. Praha 1958, 439 s.; *Aplikace matematiky na studium psaní*. Pokroky matematiky, fyziky, astronomie 3, 1958, s. 245n. a 393n. aj.

⁸¹Viz Čáp, J. – Matoušek, K. – Matula, M. – Petrásek, J.: *Frekvence slov v stenografické praxi*. Praha 1961; Matula, M.: *Frekvence kořenů slov*. Praha 1963.

⁸²Viz *Metodika vyučování těsnopisu*. Praha 1967; *Využití statistiky fonosémů pro optimalizaci těsnopisné teorie*. Zprávy SÚT, 1969/70, č. 1, s. 11n.; *Máme těsnopisnou soustavu plně vyhovující duchu českého jazyka?* Zprávy SÚT 1968/69, č. 3, s. 65n.

a vhodnější soustavy českého těsnopisu, než je soustava Heroutova-Mikulíkova, po zkušenostech se zaváděním nového systému a velkou náročností přechodu, který by trval zhruba 10 let, nebyly tyto návrhy využity.

Přestože frekvenční těsnopisné slovníky a výzkumy kladou důraz více na formu slov, než na jeho význam (slučují tedy hesla, která by v lingvisticky pojatých slovnících či seznamech byla rozdělena), je srovnání stenografických a lingvistických statistik možné i užitečné současně.

Jak již bylo uvedeno na začátku, lze těsnopisné systémy rozdělit na ruční a strojové. K nahrazení ručního stenografického písma písmem strojovým slouží *těsnopisný psací stroj*. Strojový záznam zpravidla užívá tzv. *akordů* – současného stisku více kláves. Písmo se otiskuje na úzký proužek papíru, který je navinut na cívce. Po každém úhozu se papír posune o řádku zdola nahoru. Počet kláves na stroji je co nejvíce zredukován. Práce stroje musí být nehlučná, aby nerušila okolí. Těsnopisné stroje se používaly zejména ve Spojených státech a západní Evropě (Anglie, Německo) a začaly se vyrábět i v Sovětském svazu. V USA i západní Evropě je použití těchto strojů běžné, i když dnes bývají nahrazovány počítačem, který pomocí speciálního editoru může softwarově převádět tento záznam do běžného písma. U nás začal na soustavě českého strojového těsnopisu pracovat ke konci 30. let 20. století Jaroslav Vrátný. Navrhoval klávesnici s 24 klávesami, na něž se snažil jednotlivá písmena umísťovat tak, aby to bylo ve shodě s charakterem českého jazyka. Konal proto rozsáhlá frekvenční statistická šetření českých hlásek asi na 20 000 slovech. Vypočítal, že jeho soustavou by bylo možno napsat v průměru 6 slov 7 akordy. Prakticky se výcvikem na Vrátného klávesnici zabýval Jaroslav Junek. K výrobě těsnopisných psacích strojů podle Vrátného prototypu však u nás nedošlo a dodnes nenašly těsnopisné psací stroje v rámci české těsnopisné praxe uplatnění. Převládá ruční těsnopisný zápis v kombinaci s moderní audiotechnikou. Pokusy vyvinout pro češtinu strojovou těsnopisnou soustavu zatím skončily neúspěchem.

Ačkoliv se může zdát, že těsnopis je v dnešní době v důsledku zdokonalování moderních technologií na ústupu, není to docela pravda. Aktivně se těsnopis dosud používá při veškerých jednáních Parlamentu České republiky (například při jednáních Poslanecké sněmovny Parlamentu České republiky bývá přítomno 10 stenografů). Stenografický zápis z jednání je zakotven přímo v jednacím řádu Parlamentu ČR, a to z toho důvodu, aby byly zachovány zápisy ze všech parlamentních jednání. Bohužel se těsnopis dnes již na žádné střední škole mezi povinnými předměty nevyučuje (lze si jej zvolit jako kroužek na Obchodní akademii v Dušní ulici v Praze pod vedením Heleny Janovské). Zájemci o výuku těsnopisu jsou tak odkázáni buď na samostudium z knih, nebo mohou absolvovat tříměsíční kurz Základy těsnopisu či pracovat s programem Těsnopis MINI, z nichž poslední dvě zmiňované možnosti jsou zaštitěny Státním těsnopisným ústavem.

V současnosti je těsnopisná problematika u nás náplní práce dvou institucí. První z nich je *Státní těsnopisný ústav*, který byl založen roku 1921 za účelem rozšiřování a zdokonalování oficiální těsnopisné soustavy Heroutovy-Mikulíkovy. Od roku 1999 je součástí Národního ústavu odborného vzdělávání v Praze. Koordinuje výuku předmětů technika administrativy, obchodní kore-

spondence a těsnopis na středních školách. Podílí se na mezinárodní spolupráci v rámci federace *Intersteno* (*Mezinárodní federace zpracování informací*) založené roku 1887. Naše republika je jejím členem od roku 1960 (tehdy ještě jako součást Československa). Interstemo pořádá každé dva roky mezinárodní kongresy, v jejichž rámci se koná mistrovství světa v psaní na stroji (od roku 1955) a mezinárodní soutěže v těsnopise (od roku 1961). Soutěží se vždy odděleně v jednotlivých jazykových skupinách. Prezidentem Interstema je pro dvouleté období vždy příslušník toho státu, v němž se má konat příští kongres. Vedle Státního těsnopisného ústavu v naší republice pracuje *Český těsnopisný spolek*⁸³, který byl ustaven 5. března 2001 v Praze a navázal na přerušenu činnost *Prvního pražského spolku stenografů* založeného roku 1859 na podnět Jindřicha Fügnera.

2.4 Viktor Jakovlevič Bunjakovskij

Tento významný ruský matematik zasáhl rovněž do prehistorie matematické lingvistiky. Pravděpodobně jako první totiž upozornil v roce 1847 na možnost využití matematických metod v lingvistice.



Viktor Jakovlevič Bunjakovskij se narodil 16. 12. 1804 v Baru (dnešní Ukrajina) a zemřel 12. 12. 1889 v Petrohradu (Sankt Petěrburg). Základní vzdělání získal Bunjakovskij v Moskvě v domě hraběte A. P. Tormasova, jenž byl přítelem jeho otce. Společně s druhým synem hraběte studoval Bunjakovskij od roku 1820 v cizině – nejprve v Coburgu, potom v Lausanne a nakonec v Paříži (zde studoval na Sorbonně a Collège de France, a měl tak možnost seznámit se s přednáškami například Laplace, Fouriera či Poissona). V roce 1824 získal Bunjakovskij titul

bakaláře, 1825 byl jmenován doktorem matematiky na Přírodovědecké fakultě (Collège de France). Po návratu do Ruska v roce 1826 začal vyučovat matematiku v Petrohradě na různých školách. V roce 1828 byl zvolen aspirantem Akademie věd, roku 1830 byl zvolen mimořádným členem Akademie a 1836 řádným členem Akademie věd. V letech 1846–1859 přednášel na petrohradské univerzitě (analytická mechanika, diferenciální a integrální počet či teorii pravděpodobnosti). Od roku 1864 až do doby krátce před svou smrtí byl viceprezidentem Akademie věd. Roku 1883 Bunjakovskij, který se převážně zabýval oblastí teorie čísel a teorií pravděpodobnosti, publikoval seznam svého vědeckého díla⁸⁴ čítající 108 vědeckých prací.

Roku 1847 publikoval Bunjakovskij v časopise *Sovremennik*⁸⁵ svůj článek s názvem *O možnosti vvedení opredelitelnych mer doverija k rezultatam*

⁸³Více viz na internetových stránkách www.tesnopsis.cz.

⁸⁴*Liste des travaux mathématiques de Victor Bouniakowsky*. SPb 1883.

⁸⁵Vol. 3, part II.