

Surovčák, Martin

## K využití Hammingovy metriky při výzkumu fonotaktiky novořečtiny

*Porta Balkanica*. 2015, vol. 7, iss. 2, pp. [4]-10

ISSN 1804-2449

Stable URL (handle): <https://hdl.handle.net/11222.digilib/135431>

Access Date: 29. 11. 2024

Version: 20220831

Terms of use: Digital Library of the Faculty of Arts, Masaryk University provides access to digitized documents strictly for personal use, unless otherwise specified.

# K VYUŽITÍ HAMMINGOVY METRIKY PŘI VÝZKUMU FONOTAKTIKY NOVOŘEČTINY

Martin Surovčák

Už na přelomu 20. a 30. let 20. století americký lingvista německého původu George Kingsley Zipf zkoumal relativní četnost hlásek v textu a dospěl k trojímu zjištění: a) hlásky a třídy hlásek se objevují v různých textech daného jazyka se stejnou frekvencí, b) univerzálně platí, že neznělé hlásky mají v textu přibližně dvojnásobnou frekvenci v porovnání se znělými hláskami, c) obtížněji artikulovatelné hlásky se objevují v textu s nižší frekvencí. Na základě těchto empirických pozorování poté zformuloval tzv. *první Zipfův zákon*, který říká, že součin frekvence slova (resp. hlásky) a jeho ranku je roven konstantě.\* Tuto statistickou zákonitost vysvětloval pomocí dvou protichůdných tendencí v přirozeném jazyce, a to ekonomii mluvčího (minimální úsilí při formulaci výpovědi) a ekonomii posluchače (minimální úsilí při percepci výpovědi) (ČERNÝ 1996: 253–255).

Formální obrat v lingvistice ukázal, že jistá dávka geometrické abstrakce může být v některých oblastech pro výzkum jazyka velmi podnětná, obzvláště tam, kde se lingvistika pokouší porovnávat rozdíly mezi jazykovými jednotkami nebo útvary. Myšlenka aplikovat tzv. metrický prostor<sup>1</sup> na jazyková data padla na úrodnou půdu především ve fonologii a dialektologii.

O. Šefčík (2007) popisuje abstrakci metrického prostoru ve fonologii takto: „Řetězce fonů mají svou délku, tato délka řetězce je dána počtem fonů v řetězci. Protože fony chápeme jako komponenty segmentální a diskrétní, můžeme je chápat jako minimální objekty v prostoru, které se od sebe nacházejí v různé vzdálenosti. Stejně tak i délku řetězce můžeme chápat jako vzdálenost mezi nejpravějším a nejlevějším prvkem řetězce.“ Prostor, ve kterém

<sup>1</sup> Metrický prostor představuje matematickou strukturu, pomocí níž lze formálně zavést pojem vzdálenosti. Je definován jako dvojice  $(M, \rho)$ , kde  $M$  je libovolná neprázdná množina a  $\rho$  je tzv. metrika, tedy zobrazení splňující axiom nezápornosti, totožnosti, symetrie a trojúhelníkové nerovnosti. Obecně k problematice metrických prostorů viz KOSMÁK–POTŮČEK 2004 nebo DOŠLÁ 2006.

se fony nacházejí, nazývá fonovým lineárním prostorem. Jelikož fonémy jsou rovněž jako fony diskrétní a segmentální komponenty, stejně tak tvoří uzavřenou množinu a vyskytují se v řetězcích, analogicky zavádí fonologický lineární metrický prostor a v něm rozlišuje tři různé druhy metrik (vzdáleností): *lineární, hrubou a jemnou*. Lineární metriku vztahuje na prvky v jednom celém řetězci, hrubou metriku na prvky řetězců dané délky a nakonec formuluje i jemnou metriku, o které hovoří tehdy, má-li být pomocí vzdálenosti vyjádřen rozdíl mezi jednotlivými fonémy nazývanými jako uspořádané  $n$ -tice distinktivních rysů. Právě na jemné metrice jsem postavil svou následující analýzu.

## Hypotéza

Vyjďeme ze Zipfových empirických pozorování a položme si nyní dvě základní otázky: Platí-li první Zipfův zákon (případně s Mandelbrotovými úpravami) pro relativní frekvenci slov a hlásek, lze jej vztáhnout také na kombinace fonémů (slabiky)? Jak souvisí slabičný kontrast\*\* s relativní frekvencí slabiky v textu?

V souladu jazykové ekonomie bychom mohli očekávat, že slabiky vyžadující z artikulačního hlediska větší úsilí se budou v textu vyskytovat s menší frekvencí. Nicméně zároveň platí, že příliš malý slabičný kontrast vyžaduje zvýšené percepční úsilí na straně posluchače.

## Metodologie

### Zdroj dat

Protože jsem neměl k dispozici speciální fonologický korpus<sup>2</sup> novořečtiny a Národní korpus řeckého jazyka (ETHEG) nebyl pro fonologickou analýzu vhodný, rozhodl jsem se pro účely tohoto článku vytvořit svůj vlastní referenční korpus omezeného rozsahu. Skladba textů obsažených v korpusu není nikdy nahodilá. Aby byl výsledný korpus a data z něj získaná dostatečně reprezentativní, musí splňovat několik kritérií:

- A. Rozsah** – Aby korpus pokryl i méně frekventované jevy v jazyce, musí být dostatečně obsáhlý. Oproti konvenčním korpusům pracujícím s velkým počtem gramatických a lexikálních jednotek, které musejí pojmout miliony slov, má fonotaktický korpus jednu nespornou výhodu. Fonologie totiž pracuje s jednotkami pouze v řádu několika desítek. Podle typologických průzkumů se počet segmentálních fonémů ve fonematickém inventáři pohybuje průměrně kolem 31–32 (DUBĚDA 2005: 44). Počet teoreticky možných vzájemných kombinací fonémů, tj. různých slabik, dosahuje sice řádu tisíců, ale kdyby jazykový systém využíval všechny kombinatoricky možné slabiky, byl by značně neefektivní. Navíc určité kombinace (např. sekvence tří po sobě následujících okluziv) vylučuje artikulační omezení. Reálný počet různých slabik tak dosahuje pouze několika stovek (např. v češtině zhruba kolem 700, viz BIČAN 2013). Z toho vyplývá, že fonotaktická analýza si vystačí s daleko menším rozsahem, než je u konvenčních korpusů obvyklé.
- B. Aktuálnost vstupních dat** – Texty obsažené v korpusu by měly být pokud možno přibližně stejné datace, a protože poměrně rychle zastarávají, měly by být také co nejaktuálnější, ideálně ne starší než pět let. V případě fonologického výzkumu není nutné se držet tohoto kritéria příliš striktně, protože fonologický systém jazyka podléhá změnám v čase daleko pomaleji než morfologie a slovní zásoba.

<sup>2</sup> Jazyková data jsou vždy zprostředkována, jelikož reprezentují jazykový systém, který sám o sobě bezprostředně přístupný není, a o němž vypovídají pouze částečně. Jazyková data jsou v promluvě organizována lineárně a jejich zpětným dekódováním lze dospět k vícedimenzionálním strukturám jazykového systému. Při analýze izolovaných textů pak logicky nastává problém chybějícího kontextu, který specifikuje informaci potřebnou pro analýzu významu a funkce. Materiálem a zdrojem jazykové analýzy jsou tedy texty, ať již psané nebo mluvené, které by měly splňovat požadavky na reprezentativnost a věrnost (tj. výběr autentických, neupravených vzorků) (ČERMÁK 1993: 3–7). V tomto směru se v posledních dvou desetiletích osvědčily jazykové korpusy, které především eliminují riziko předpojatého přístupu lingvisty ke zkoumanému materiálu a díky tomu představují nejobhatší a nejrealističtější zdroj poznání jazyka vůbec. Jazykovým korpusem se rozumí „vnitřně strukturovaný, unifikovaný a obvykle i oindexovaný a ucelený rozsáhlý soubor elektronicky uložených a zpracovávaných jazykových dat většinou v textové podobě, organizovaný se zřetelem k využití pro určitý cíl, vůči němuž pak je také považován za reprezentativní“ (ČERMÁK 1995: 119).

\* Francouzský matematik B. Mandelbrot později poukázal na to, že první Zipfův zákon sice v hrubých rysech odává spád závislostní křivky mezi frekvencí hlásky a jejím rankem, přesto nevystihuje dostatečně podrobně její průběh (viz SEDLAČIKOVÁ 2012: 64–65). Mandelbrot proto upravil původní Zipfův vztah jako tzv. *kanonický zákon*:

$$p_r = P(r + \rho)^{-\beta}$$

kde

- $p_r$  je frekvence příslušného slova  
 $\beta$  je konstanta pro daný text  
 $\rho$  je konstanta korigující frekvenci slov s nízkým pořadím  
 $P$  je konstanta (charakteristika rozsahu výběru)  
 $r$  je rank nabývajících pro každou jednotku jiné hodnoty

\*\* Český fonolog B. Hála formuloval myšlenku, že podstatou slabiky je přechod od strikturní k apertuře a toto tvrzení klade do souvislosti s procesem glotogeneze. Jednotlivé hlásky totiž podle něj vznikly patrně diferenciací a stabilizací částí slabiky teprve po oddělení konsonantické přetruhy od vokálního jádra. Více fyziologicky motivovaný názor pak zastává P. MacNeilage, v jehož teorii „rámce a obsahu“ hybný rámeček slabiky představuje tzv. mandibulární cyklus nutný zejména pro žvýkání potravy. Během evoluce byl tento rámeček postupně korelovan s fonačními a artikulačními vzorci, jež se následně staly jeho obsahem. Za nejpřirozenější obsah rámce autor této teorie pokládá sekvenci bilabálního konsonantu a středního vokálu, k jejíž artikulaci stačí pouhý pohyb čelisti (DUBĚDA 2005: 136).

### Mgr. et Mgr. Martin Surovčák (1988)

Vystudoval balkanistiku a neogrecistiku na Masarykově univerzitě v Brně, kde v současné době pokračuje v doktorském studiu neogrecistiky. Uveřejněný článek je součástí jeho dizertačního projektu. Působí jako akademický pracovník Ústavu řeckých a latinských studií Filozofické fakulty Karlovy univerzity v Praze. Rovněž pracuje jako překladatel a tlumočnick albanštiny a novořečtiny na volné noze.

[martin.surovcak@ff.cuni.cz](mailto:martin.surovcak@ff.cuni.cz)

Filozofická fakulta Univerzity Karlovy, Ústav řeckých a latinských studií, Palachovo nám. 2, Praha 1

**C. Zdroje různé povahy** – Vzhledem k tomu, že jak psaný, tak i mluvený projev mají svá zvláštní specifika, měl by korpus ideálně obsahovat psané texty i přepisy mluvených textů. Mluvené projevy, zvláště v neformálních komunikačních situacích, se od těch psaných liší nejen např. specifickou slovní zásobou, ale i tím, že v nich ve větší míře probíhají fonologické alternace, kterým se psaný jazyk často záměrně vyhýbá (srov. elize, apokopa, disimilace skupin  $\pi\tau$ ,  $\phi\theta$ ,  $\chi\theta$  v novořečtině). Přepis mluvených projevů však představuje časově i technicky náročný proces, a proto jsem pro potřeby tohoto článku výzkum omezil pouze na psaný jazyk.

**D. Stylistická různorodost** – Korpus by měl být stylisticky vyvážený. Měl by v „ideálním“ poměru zahrnovat beletrii, publicistický, vědecký a administrativní styl, nicméně ideální poměr není nijak univerzálně daný. Např. Britský národní korpus obsahuje informativní a imaginativní texty v poměru 2:3 (viz ČERMÁK 1995: 125). V Řeckém národním korpusu (ETHEG) je zastoupen téměř ze dvou třetin denní tisk (61,29 %), následují knihy (9,41 %), časopisy (5,89 %), internet (0,32 %) a zbytek (23,08 %) tvoří jiné zdroje.<sup>3</sup>

S přihlédnutím k výše uvedenému jsem jako referenční text zvolil první kapitolu z historického románu *Έρωσ, Θέρος, Πόλεμος* autorky Evjenie Fakinosové (Atény, 2003). Text je napsán jazykem, který se prakticky nevzdaluje od spisovného standardu, až na několik málo výjimek nezabíhá do katharevusy (archaizujícího stylu novořečtiny) a obsahuje málo vlastních a místních jmen, číselných údajů, cizojazyčných vsuvek apod.

### Fonologická transkripce výchozího textu

Výchozí text je v této fázi zapotřebí převést do podoby fonologického zápisu a následně ho sylabifikovat (rozčlenit na slabiky).

Pro spisovnou novořečtinu předpokládám existenci 5 vokálních fonémů:

/a/, /e/, /i/, /o/, /u/

a 19 konsonantických fonémů:<sup>4</sup>

<sup>3</sup> Viz oficiální stránky Řeckého národního korpusu <<http://hnc.ilsp.gr/info.asp>>.

<sup>4</sup> Srov. SETATOS (1974), PETROUNIAS (2002: 461), NESPOR (1999: 69), ARVANITI (1999). Univerzální shoda nicméně neexistuje. Diskuzi vzbuzuje především otázka statu-

/p/, /b/, /f/, /v/, /m/, /t/, /d/, /θ/, /ð/, /s/, /z/, /n/, /l/, /r/, /j/, /k/, /g/, /x/, /ɣ/,

které jsou rozlišitelné 13 binárními distinktivními rysy:

VOKALICKÝ, KONSONANTICKÝ, NÍZKÝ, VYSOKÝ, PŘEDNÍ, ZADNÍ, NAPJATÝ, LABIÁLNÍ, NAZÁLNÍ, KONTINUÁLNÍ, VIBRANTNÍ, DENTÁLNÍ, ZNĚLÝ.<sup>5</sup>

Sylabifikace<sup>6</sup> (rozdělení slova na slabiky) je nezávislá na morfologii i na samotném počtu segmentů, z nichž se slabika skládá. Řídí se dvěma principy: a) *principem sonority*, b) *principem maximální iniciály*. Princip sonority je postaven na skutečnosti, že míra sonority jednotlivých segmentů dosahuje maxima na slabičném vrcholu, zatímco směrem ke svahům slabiky klesá, eventuálně po dosažení maxima zůstává stejná (neklesá). V hierarchii sonority jsou přitom hlásky seřazeny takto: nízké vokály (maximální sonorita), vysoké vokály / glidy, likvidy, nazály, frikativy, okluzivy (minimální sonorita). Princip maximální iniciály říká, že konsonanty v kontextu VCV jsou v souladu s univerzálními a/nebo jazykově specifickými požadavky na fonotaktiku slabik sylabifikovány jako slabičné iniciály (ZIKOVÁ 2014: 37). Např.: slovo „άστρο” musí být sylabifikováno jako /a.stro/, nikoliv /as.tro/, jelikož fonotaktika novořečtiny připouští na začátku slova sekvenci /str/, srov. στρώνω, στρέμμα, στραβός apod.<sup>7</sup>

Fonologický přepis a sylabifikaci textu jsem provedl zčásti automaticky pomocí funkce nahrazování znaků a případné nesrovnalosti jsem odstranil manuálně. Výsledek transkripce vypadá následovně, např.:

su afrikát, glidy (viz např. SOULTATIS 2013) a palatál. Znělé okluzivy /b/, /d/, /g/ jsou sice přijímány jako samostatné fonémy, ale jejich distribuce je omezena prakticky jen na slova cizího původu (romanismy, turcismy, anglicismy apod.).

<sup>5</sup> Pořadí distinktivních rysů při převodu fonémů na binární kódy je závazné, byť arbitrární.

<sup>6</sup> V zásadě existují dva přístupy ke slabičné struktuře: *derivativní* a *reprezentační*. Derivační přístup předpokládá, že slabičná struktura slova je odvoditelná z lineárního uspořádání jeho segmentů, že každý sonoritní vrchol v řetězci segmentů projektuje slabičné jádro a že hranice slov korespondují s hranicemi slabik. Reprezenční přístup naopak vychází z předpokladu, že slabičná struktura slova je principiálně nezávislá na segmentálních řetězcích, dále že ne každá slabika musí nutně projektovat sonoritní vrchol (tzn. připouští existenci prázdných slabičných jader) a že hranice slov nemusejí nutně korespondovat s hranicemi slabik (ZIKOVÁ 2014: 6).

<sup>7</sup> Alternativní pojetí sylabifikace v novořečtině v souladu s teorií optimality nabízí např. KAPPA (1997).

Ένα άδειο πουκάμισο φιδιού. Το είδε που ξάσπριζε κάτω από τα σκοινιά.

||e.na#a.ðjo#pu.ka.mi.so#fi.ðju||to#i.ðe#pu#ksa.spri.ze#ka.to#ap#ta#ski.na||

Celkem se podařilo ve výchozím textu izolovat 6 619 výskytů 477 různých slabik.

### Hammingova vzdálenost

Nyní se vraťme zpět ke slabičnému kontrastu, jak byl předeslán v hypotéze. Kontrast mezi vrcholem a svahy slabiky představuje jedno ze základních hledisek delimitace slabik. Řeč si můžeme představit jako souvislý informační tok, kdy je mluvčí při procesu kódování neustále nucen mezi jednotlivými segmenty slabiky „přepínat“ hodnoty distinktivních rysů. Čím vícekrát toto „přepnutí“ proběhne v rámci jedné slabiky a čím více změn distinktivních rysů je uplatněno, tím je i z artikulačního hlediska tato realizace slabiky pro mluvčího náročnější. Na straně posluchače je situace přesně opačná. Markantnější rozdíly totiž usnadňují proces percepce.

Pokud přepíšeme fonémy do podoby binárních kódů, můžeme slabičný kontrast vyjádřit pomocí tzv. *Hammingovy vzdálenosti*<sup>8</sup>, která je definována jako nejmenší počet pozic, na kterých se řetězce stejné délky daného kódu liší, jinými slovy počet

záměn, které je potřeba provést pro změnu jednoho z řetězců na druhý (HAMMING 1950). Jednotkou Hammingovy vzdálenosti je *bit*.

Vezměme si jako modelový příklad dvě otevřené slabiky: /ma/ a /lu/. Rozložíme-li tyto fonémy na svazky distinktivních rysů, přičemž délka řetězce bude pro každý foném stejná a pořadí distinktivních rysů závazé, kladná hodnota bude vyjádřena 1 a záporná 0, pak dostaneme tento výsledek:

/m/ [1101010110001]  
/a/ [1010001001001]  
/l/ [1101100001001]  
/u/ [1001011001001]

Zvýrazněny jsou distinktivní rysy, v nichž se uvedené fonémy liší. Vidíme tedy, že Hammingova vzdálenost v případě slabiky /ma/ je rovna 8 bitům a v případě slabiky /lu/ má hodnotu 4 bity. Slabičný kontrast tedy logicky roste s délkou slabik, protože Hammingovy vzdálenosti mezi jednotlivými fonémy se navzájem sčítají. Např. u slabiky /mam/ se jejich součet rovná 16 bitům.

Speciální případ nastane u slabik, které jsou tvořeny pouze holým vokalickým jádrem (tzn. neobsahují ani iniciálu, ani kódu). Z logiky věci vyplývá, že hodnota Hammingovy vzdálenosti v tomto případě nemůže být určena.

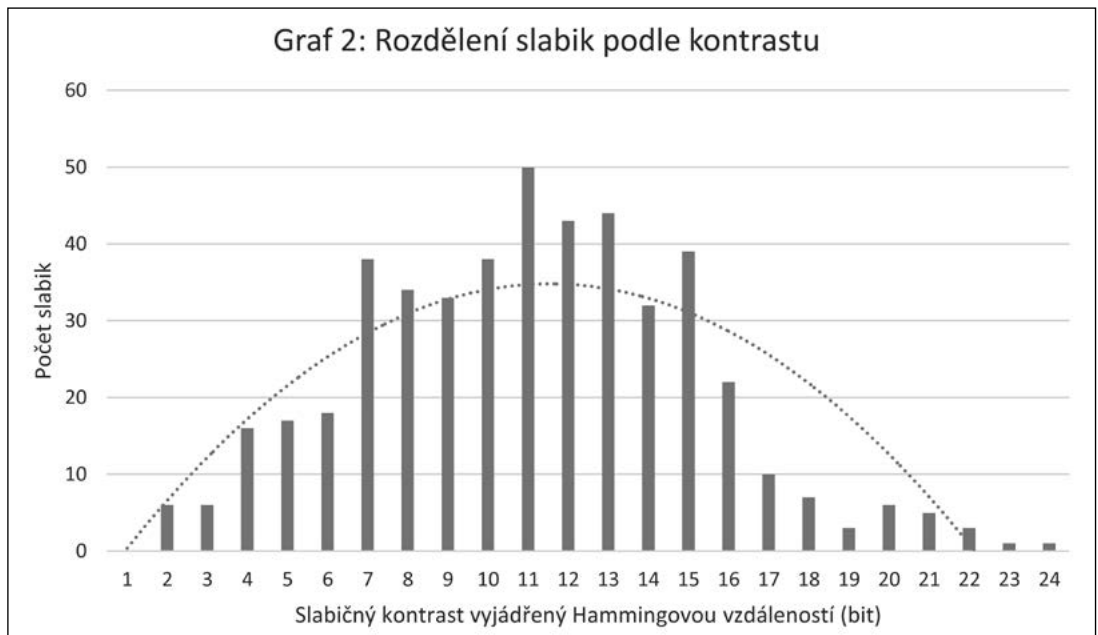
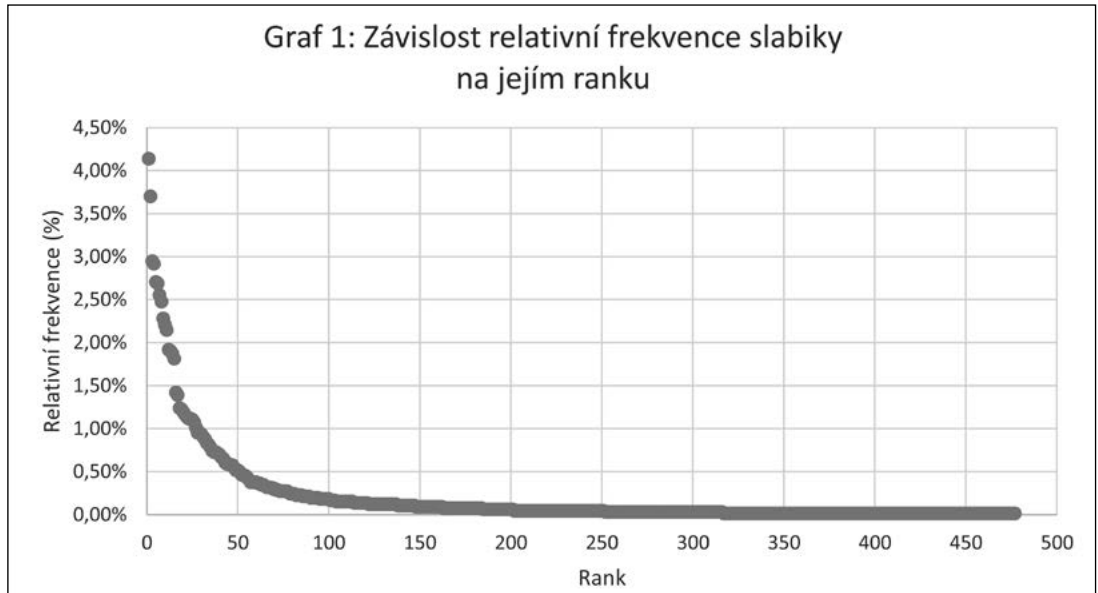
### Výsledky analýzy

Kvůli omezenému prostoru zde nemůže být publikován kompletní výčet všech 477 identifikovaných slabik, proto uvádím pouze několik modelových příkladů:

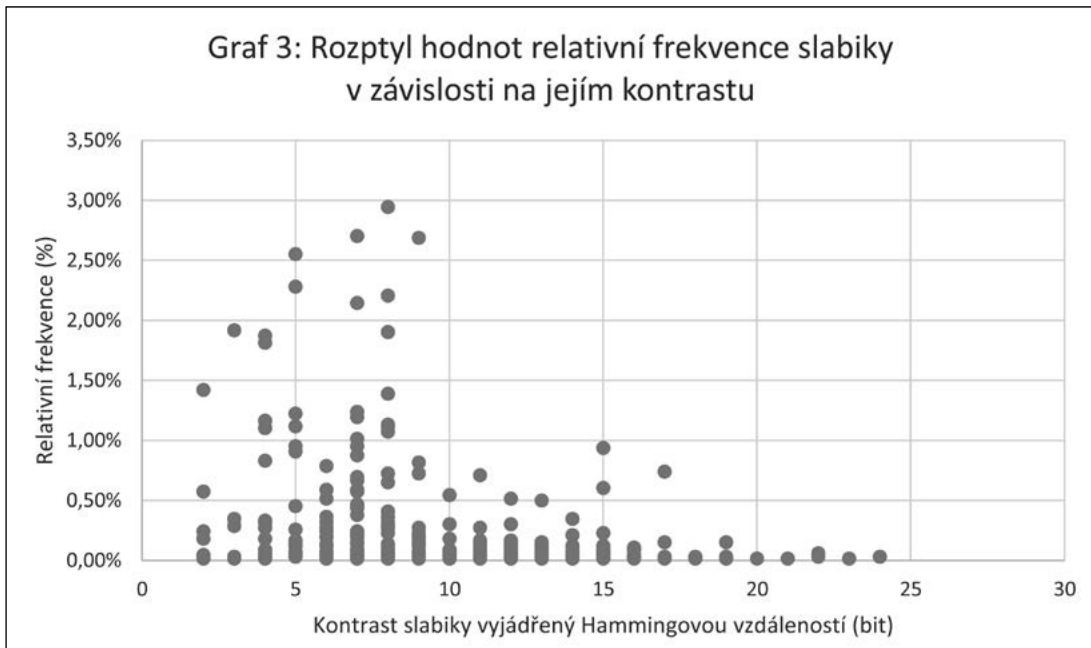
Rank	Slabika	Typ	Relativní frekvence (%)	Slabičný kontrast (bit)
2.	/e/	V	3,70 %	–
4.	/a/	V	2,92 %	–
21.	/ne/	C <sub>son</sub> V	1,16 %	4
43.	/pro/	C <sub>okl</sub> C <sub>vibr</sub> V	0,60 %	15
58.	/lja/	C <sub>son</sub> C <sub>kon</sub> V	0,38 %	7
77.	/xan/	C <sub>kon</sub> VC <sub>son</sub>	0,27 %	11
91.	/as/	VC <sub>sib</sub>	0,20 %	6
134.	/xti/	C <sub>kon</sub> C <sub>okl</sub> V	0,12 %	11
196.	/smos/	C <sub>sib</sub> C <sub>son</sub> VC <sub>sib</sub>	0,06 %	22
231.	/spri/	C <sub>sib</sub> C <sub>okl</sub> C <sub>vibr</sub> V	0,05 %	15

<sup>8</sup> Richard Wesley Hamming (1915–1998) byl americký matematik, jehož objevy přispěly především k rozvoji informatiky a telekomunikačních technologií.

Seřadíme-li hodnoty relativní frekvence slabik sestupně, můžeme pozorovat, že hodnoty zanesené do grafu č. 1 celkem věrně sledují křivku nepřímé úměrnosti, což dokazuje platnost prvního Zipfova zákona i pro slabiky.



Nyní se zaměříme na druhý bod hypotézy. Zjišťujeme, že rozpětí hodnot slabičného kontrastu vyjádřeného Hammingovou vzdáleností se pohybuje od minimální hodnoty 2 bitů (např. /li/, /le/ aj.) až po maximální hodnotu 24 bitů, která byla zjištěna u slabiky /fron/. Zároveň můžeme pozorovat, že s rostoucí Hammingovou vzdáleností narůstá i počet slabik, které příslušná hodnota Hammingovy vzdálenosti charakterizuje, ale od určitého bodu začíná opět poměrně rychle klesat. Zjištěné hodnoty slabičného kontrastu v podstatě kopírují Gaussovu křivku, z čehož můžeme usoudit, že její vrchol představuje optimální hodnotu slabičného kontrastu. Pro novořečtinu byla na základě analýzy zvoleného materiálu zjištěna optimální hodnota 11 bitů.



Trochu jiný obraz získáme zohledněním relativní frekvence slabik. V grafu č. 3 je každá slabika zobrazena jako bod. Nyní lze pozorovat, že slabiky s nejvyššími hodnotami relativní frekvence v textu jsou soustředěny v pásmu, které odpovídá nižším hodnotám slabičného kontrastu (zhruba 6–8 bitů), než je zjištěná optimální hodnota. Velký podíl na tom mají jednoslabičná slova, která se v textu mnohokrát opakují a tím zvyšují i relativní frekvenci dané slabiky. Zde je několik nejčastěji zastoupených:

Rank	Slabika	Význam	Relativní frekvence (%)	Slabičný kontrast (bit)
1.	/i/	η – člen urč. nom. sg. f. οι – člen urč. nom. pl. m. a f. ή – spojka vylučovací	4,14 %	–
3.	/ke/	και – slučovací spojka	2,95 %	8
5.	/na/	να – částice konjunktivu	2,70 %	7
6.	/to/	το – člen urč. nom. sg. n. a ak. sg. m./n.	2,69 %	9
7.	/ti/	τη – člen urč. ak. sg. f. τι – tázací zájmeno	2,55 %	5
8.	/o/	ο – člen urč. nom. sg. m.	2,48 %	–
11.	/me/	με – předložka	2,14 %	7
13.	/tis/	τις – člen urč. ak. pl. f.	1,90 %	8
18.	/tu/	του – člen urč. gen. sg. m./n.	1,24 %	7

## Závěry

1. Analýzou zvoleného jazykového materiálu jsem došel k závěru, že první Zipfův zákon platí i pro slabiky jako fonémové kombinace.
2. Slabičný kontrast jsem vyjádřil pomocí součtu Hammingových vzdáleností mezi foné-

my, z nichž se slabika skládá. Zjistil jsem, že v souladu s teorií optimality existuje optimální hodnota slabičného kontrastu, která je v novorečtině rovna 11 bitům. Slabiky s větším kontrastem jsou sice snáze perceptovatelné posluchačem, ale vyžadují větší artikulační úsilí

na straně mluvčího. Naopak slabiky s menším kontrastem jsou výhodné z hlediska ekonomie mluvčího, ale znesnadňují proces percepcí na straně posluchače.

3. Dále jsem zjistil, že frekventovaná jednoslabičná slova mají často nižší než optimální hodnotu slabičného kontrastu. To lze interpretovat tak, že tyto slabiky jsou „energeticky úsporné“ pro mluvčího, ale na druhou stranu příliš nezatěžují z hlediska percepcí ani posluchače právě proto, že se v textu mnohokrát opakují.
4. Fonotaktika novořečtiny připouští maximálně čtyřčlenné kombinace fonémů v rámci slabiky. Do uvedené optimální vzdálenosti 11 bitů se „vešly“ tříčlenné kombinace typu CVC, dokonce i několik čtyřčlenných kombinací typu CCVC. Je tedy patrné, že u vícečlenných kombinací musí být artikulační pohyb plynulý, aby na úkor efektivity zbytečně nenarůstala hodnota slabičného kontrastu. ■

## Bibliografie

### Prameny:

FAKINO, E. (2003) *Eros, Theros, Polemos*. Athina: Ekdoseis Kastanioti.

### Sekundární literatura:

ARVANITI, A. (1999) Standard Modern Greek. *Journal of the International Phonetic Association* 29, vol. 2, 167–172.

BIČAN, A. (2013) *Phonotactics of Czech*. Frankfurt am Main: PL Academic Research.

ČERMÁK, F. (1993) *Základy lingvistické metodologie*. Praha: Karolinum.

ČERMÁK, F. (1995) Jazykový korpus: Prostředek a zdroj poznání. *Slovo a slovesnost*. 1993, roč. 56, 119–134.

ČERNÝ, J. (1996) *Dějiny lingvistiky*. Olomouc: Votobia.

DOŠLÁ, Z. (2006) *Metrické prostory: teorie a příklady*. Brno: Masarykova univerzita.

DUBĚDA, T. (2005) *Jazyky a jejich zvuky*. Praha: Karolinum.

HÁLA, B. (1956) *Slabika, její podstata a vývoj*. Praha: Československá akademie věd.

HAMMING, R. W. (1950) Error Detecting and Error Correcting Codes. *The Bell System Technical Journal*. Vol. XXIX, No. 2.

KAPPA, I. (1997) Syllaviki domi tis Neas Elliniki sti theoria tou Veltistou. *Meletes gia tin elliniki glossa* 17, 1997, 83–91.

KOSMÁK, L. – POTŮČEK, R. (2004) *Metrické prostory*. Praha: Academia.

NESPOR, M. (1999) *Fonologia. Prosarmogi stin elliniki glossa*. Athina: Patakis.

PETROUNIAS, E. (2002) *Neoelliniki grammatiki kai synkritiki («antiparathetiki») analysi. Tomos 1: Fonitiki kai eisagogi sti fonologia*. Thessaloniki: Ekdoseis ZITI.

SETATOS, M. (1974) *Fonologia tis Koinis Neoellinikis*. Athina: Papazis.

SOULTATIS, Th. (2013) *The status of the glide in Modern Greek*. Dostupné z <[https://www.ucl.ac.uk/pals/research/linguistics/publications/wpl/13papers/Soultatis\\_UCLWPL2013.pdf](https://www.ucl.ac.uk/pals/research/linguistics/publications/wpl/13papers/Soultatis_UCLWPL2013.pdf)> .

ŠEFČÍK, O. (2006) Fonologická vzdálenost a přednost mezi termíny opozice. In: *Teorie a empirie. Bichla pro Krčmovó*. Brno: Masarykova univerzita, 297–304.

ŠEFČÍK, O. (2007) K užití metriky ve fonologii. *Sborník prací Filozofické fakulty brněnské univerzity. A, Řada jazykovědná = Linguistica Brunensia*. 2007, roč. 56, č. A55, [19]–25.

ZIKOVÁ, M. (2014) *Úvod do slabičného typologie*. Brno: Masarykova univerzita.

## SUMMARY & KEYWORDS

### On the Use of the Hamming Metric in Research into the Phonotactics of Modern Greek

The article presented here provides evidence that Zipf's First Law applies for syllables in Modern Greek and introduces the usage of the Hamming distance to describe the dependence between syllables' contrast and their relative frequency within a text. A custom corpus was used as the source of linguistic data.

► phonotactics; phonostatistics; Modern Greek; Zipf's law; Hamming metric

Předkládaný článek dokládá platnost tzv. *prvního Zipfova zákona* pro novořečké slabiky a představuje využití Hammingovy vzdálenosti při popisu vzájemné závislosti mezi kontrastem slabiky a její relativní frekvencí v textu. Jako zdroj jazykových dat byl použit vlastní korpus.

► fonotaktika; fonostatistika; novořečtina; Zipfův zákon; Hammingova metrika