

Dan Faltýnek, Vladimír Matlach

Gramatiky DNA

Edice Qfwfq

Olomouc
2014

Gramatiky DNA: Báze nejsou písmena

Dan Faltýnek, Vladimír Matlach

Recenzovali:

Prof. PhDr. Jan Kořenský, DrSc.

Prof. RNDr. dr hab. Jan Andres, DSc.

Edice Qfwfq, sv. 18

Tato publikace vychází v rámci grantu Inovace studia obecné jazykovědy a teorie komunikace ve spolupráci s přírodními vědami.

Reg. č. CZ.1.07/2.2.00/28.0076.

Tento projekt je spolufinancován Evropským sociálním fondem a státním rozpočtem České republiky.

Neoprávněné užití tohoto díla je porušením autorských práv a může zakládat občanskoprávní, správněprávní, popř. trestněprávní odpovědnost.

1. vydání

© Dan Faltýnek, Vladimír Matlach, 2014

© Univerzita Palackého v Olomouci, 2014

ISBN 978-80-244-4341-6

Obsah

Předmluva	5
Hlásky	9
Hlásková intuice	9
Lingvistický popis hlásky	11
Analogie hlásky v genetickém kódu	20
Báze nejsou písmena	31
Genetický kód	31
Genetický kód a přirozený jazyk	33
Instrumentárium molekulární biologie	34
Lingvistická paralela	38
Motivace experimentu	39
Metodika kvantitativní analýzy DNA	39
Zipfův zákon na přirozených textech	43
Ověření metafory DNA	49
Aplikace metody	49
Interpretace výsledků experimentu	53
Konstruování gramatik DNA	55
Co je to gramatika	55
Jakobsonovy analogie jazyka a DNA	58
Tvorba gramatiky	61
Tvorba gramatiky DNA od výrazu a od obsahu	66
Konstrukce gramatiky ze sémantické báze	68

Závěr	73
Literatura	75
Souhrn	85
Summary	87
Přílohy	89
Rejstřík věcný	117
Rejstřík jmenný	121

Předmluva

Záměrem této knihy je zorientovat čtenáře v různých pojetích genetického kódu a také předložit naše vlastní pojetí. Na následujících stránkách reflektujeme tradiční pojetí genetického kódu, který od původu reprezentuje tabulka bází DNA a k nim přiřazovaných aminokyselin. Chceme čtenáři ukázat, jak se představa genetického kódu rodila u objevitelů struktury DNA a molekulárních biologů, kteří s touto myšlenkou začali pracovat již na přelomu padesátých a šedesátých let dvacátého století, ale také to, jak s koncepcí genetického kódu pracuje moderní molekulární genetika. Od tabulky genetického kódu postoupilo uvažování nad genetickým kódem mnohem dál. Dnešní molekulární genetika je prostoupená termíny analogizujícími buněčné procesy a přirozený jazyk. Zkušenost s jazykem pomáhá molekulárním genetickým při orientaci ve změti interakcí vedoucích od genetické informace k vlastnostem organismů. Humanitní vědci se naopak ptají na to, zda je skutečně možné vidět v tématech biologie jazykové znaky, které jsme jinak zvyklí spojovat jen s činností člověka. Sémiotika upřela jazyk zvířatům, protože u nich nenašla dostatečnou komplexnost symbolické reprezentace, kterou od jazyka očekáváme. Oproti tomu mnohem níž, na buněčné úrovni, se vědecké uvažování dopracovává k dalším a dalším jazykovým vlastnostem. V tomto směru uvažování se zcela ztratil myšlenkový rozměr, který byl nejspíš nejdůležitějším, i když implicitním, důvodem „odejmutí“ jazyka zvířatům. U buněčných procesů s žádnou myšlenkovou činností přirozeně nepočítáme. Přesto ale ve vědeckém poznání života přibývá domněnek o blízkosti jazyka a jevů, které jsou podstatou fungování organismů. Proč? Navzdory rozdílu mentální reprezentace jazyka a fyzické povahy života se objevuje stále větší podobnost jejich fungování. Život a jazyk mají podobný systém.

V této knize jsme utřídili naše zkoumání znakových procesů v buňce. Po delší době jsme dospěli k názoru, že pokud se vůbec takové procesy v buňce objevují, tak bychom při jejich popisu měli trvat na několika základních předpokladech. Z těch vyplynula překvapivě

soudržná podoba genetického kódu. Ta bohužel neodpovídá tomu, jak o genetickém kódu mluví molekulární biologie, ani tomu, jak popisuje znakové procesy v buňce biosémiotika, věda mající ve svém programu propojení popisu jazyka a života. O to zajímavější je úkol vysvětlit naše závěry na následujících stranách.

Naše nové pojetí genetického kódu představíme ve třech oddílech. Výklad knihy začíná detailní rozpravou nad tím, jakou pozici mají v přirozených jazycích hlásky. Hlásky jsou jednou z klíčových vlastností kódu, každý přirozený jazyk má svůj soubor hlásek. Proto se u tohoto sémiotického předpokladu zastavujeme a popisujeme, na základě jakých vlastností tvoří hlásky hláskový systém, jak se mění jejich vlastnosti v kombinaci s jinými hláskami ve slově, jak velký vliv mají na hláskový systém nejrůznější ekonomizační tendence a další jazykové roviny, jako je rovina slova či morfému. Může to být neintuitivní, ale i oněch pár hlásek jazyka tvoří překvapivě důmyslný komunikační aparát.

Hlásky, respektive písmena, která hlásky reprezentují v psaném textu, jsou součástí původní analogie jazyka a DNA. To je hlavní důvod toho, proč jim v této knize věnujeme takovou pozornost. V tabulce genetického kódu nacházíme báze DNA, které v trojkombinacích odpovídají vždy jedné aminokyselině přidělované do proteinového řetězce při proteosyntéze. Báze jsou tradičně analogizovány s písmeny textu přirozeného jazyka a právě pojetí bází DNA jakožto písmen či hlásek genetického kódu se v této knize snažíme odmítnout. Tomuto tématu je věnována první kapitola, ve které předkládáme argumenty založené na užití základních sémiotických předpokladů, jako je arbitrárnost znaku, dvojí artikulace kódu atd. Ve druhé kapitole popisujeme experiment, který naše domněnky o povaze hlásky v genetickém textu potvrzují. Použili jsme pro to metodu zkoumání projevů Zipfova zákona na n-gramech genetického textu. Ukazujeme také historii kvantitativních analýz DNA a snažíme se dát náš nový přístup ke kvantitativní analýze genetického textu do souvislosti s předešlými výzkumy. Výsledky analýz jsou uvedeny na konci druhé kapitoly. Zároveň jsme do přílohy knihy umístili množství grafů, na nichž můžeme výsledky naší metody sledovat.

Závěrečná kapitola knihy ukazuje to, co znamená, když nemluvíme jen o genetickém kódu, ale navíc i o gramatice DNA. Pod pojmem gramatika rozumíme většinou gramatiku nějakého národního jazyka. Chceme ukázat, že stejně tak dobře je gramatikou každý systémový popis textu nebo znakového chování. Vycházíme zde ze závěrů, které jsme učinili při analýze hláskové roviny genetického textu, a ukazujeme, jak na jejich základě můžeme postupovat při tvorbě gramatiky. V teorii gramatiky se rozlišují dva protipólné přístupy. První z nich popisuje jazyk na základě analýzy jeho výrazových prostředků, což znamená, že se zaměřuje na to, co jednotlivé znaky reprezentuje – hlásky, slova, jejich části a tvary, postavení ve větě atd. Druhý přístup vychází z analýzy toho, co jednotlivé jednotky textu vyjadřují a jaký mají význam. Ačkoliv se s významem nepracuje snadno, protože jej nemáme jako text bezprostředně k dispozici, ukazuje se cesta od významu jako výhodnější způsob tvorby gramatiky, protože je mnohem odolnější vůči chybám při přidělování významu určité části textu. Význam je totiž často delegován mezi více jednotek a výrazový přístup jej má tendenci přidělit první, která se nabízí, další nechává v gramatickém popisu zapomenuté. Přístup od významu pak zkusíme navrhnout jako nejvhodnější při tvorbě gramatik DNA – inspirací je nám pro to způsob konstrukce gramatiky ze sémantické báze, který navrhl Jan Kořenský.

Tuto knihu jsme se snažili napsat tak, aby byla co nejpřístupnější. Spojení lingvistiky, biologie a kvantitativních metod se každopádně může jevit poněkud nesourodě. Gramatický popis něčeho, co je fyzické a živé – to je představa, která může vyvolávat rozpačitý dojem. Na druhou stranu se výzkum fyzického a živého odvíjí na základě analýz genetických textů v podobě posloupností písmen A, C, T, G, a právě hledání vztahů mezi částmi textu, nalézání a popisování jeho gramatiky, je zájem lingvistiky.

Hlásky

Hlásková intuice

Hlásky jsou jazykové jednotky, které člověk identifikuje intuitivně. Činí tak, aniž by při tom musel mít předem popsány vlastnosti, kterými se jedna hláska odlišuje od druhé. K rozpoznání jednotlivých hlásek nemusí člověk znát ani vytvářet jejich kompletní seznam – to po něm vyžaduje až školní praxe. Memorování abecedy nebo deklamace správného pořadí jejích členů tvoří kolorit prvních kroků školským vzdělávacím systémem. Abeceda je velice silná instituce, která organizuje každou kulturu založenou na psaných textech, počínaje seznamy textů, lidí a věcí a konče zkratkami jejich názvů a jmen. Znalost pořadí grafémů abecedy odráží zapojení člověka do psané kultury, hlásku ale umí z řetězce textu vyčlenit i bez toho.

Oproti vymezení jednotlivé hlásky není popis vlastností, které tyto hlásky odlišují, nijak intuitivní. Takový popis vyžaduje, aby se na soubor hlásek pohlíželo systematicky a aby byl při tom použit popisný aparát vycházející z analýzy textu nebo z precizních experimentálních metod (některé z nich provádí např. profesor Higgins z filmu *My Fair Lady*, jehož předobrazem byl anglický fonetik Henry Sweet). Každý záměr seřadit všechny hlásky jazyka a popsat jednotným způsobem jejich rozdíly je motivován vůlí nahlédnout za nevědomé používání řečových schopností a je také něčím, co řeč sama o sobě nijak nepotřebuje a nevyžaduje. Popsat, čím se od sebe hlásky odlišují, není zcela samozřejmá činnost, protože je nutno obrátit jazyk, nástroj poznávání, na sebe sama. Taková reflexivní činnost spočívá v procedurách, které v naší kultuře a způsobu života většinou náleží až vědě. Na druhou stranu, už starověké indické či latinské mluvnice uměly zcela přijatelně, na základě artikulačních a částečně akustických vlastností, třídit hlásky na skupiny se společnou vlastností. Pokud tedy není lidská snaha poznat repertoár všech hlásek a jejich rozdíly dílem intuice, je to jistě přirozený projev lidských poznávacích aktivit, kterému navíc v historii vždy

příslušelo oceňované postavení (příkladem může být židovská mystika, tzv. kabala a další).

Ve vývoji dětské řeči jsou hlásky rozpoznávány až po stádiu, kdy se dítě zaměřuje na slabiky. Slabiku lze definovat různě (viz Hála 1956), jako pravidelnou kombinaci hlásek vázanou na zaznění tónu, ale také jako nejmenší možný artikulovatelný úsek řeči. Hlásky jsou samostatně, mimo slabiku, nevyslovitelné, nebo jejich samostatná realizace neodpovídá realizaci ve slabice. Za samostatně vysloveným konsonantem vždy následuje vokál nebo alespoň vokál neutrální (volné zaznění hlasu, tzv. hezitační zvuk). Pokud bez následného vokálu nebo hezitačního zvuku vyslovíme hlásku /p/, bude se námi vytvořený zvuk podobat prasknutí bubliny, ale určitě ne hlásce /p/. Když hlásku pojmenováváme, například při výčtu písmen abecedy, napomáháme si dodáním vokálu, v češtině tradičně [é], v angličtině [í]. Samostatné vyslovení vokálů se naopak jeví jako samozřejmé. Vokál ale v takovém případě nevykazuje délku, která je při jeho zapojení do slabik a slov vždy dodržována (délka vokálu je navíc závislá na jeho postavení ve slově / přízvukovém taktu, např. vokál na konci slova se liší délkou od stejného vokálu uvnitř slova). Pro vokály je délka tzv. distinktivní vlastností, to znamená, že odlišuje dlouhé a krátké vokály, respektive slova, ve kterých se vyskytují (např. dál a dal). To, že jsou hlásky samostatně nevyslovitelné, by mohlo odporovat tvrzení, že jednotlivé hlásky vydělujeme intuitivně. Tak tomu ale není. Hlásky jsou identifikovány jako ty části slabik, které se v jednotlivých slabikách pravidelně opakují a z nichž se slabiky vždy tvoří.

Za historicky první systematickou analýzu souboru hlásek určitého jazyka můžeme považovat vytvoření hláskového písma. Většina vyspělých kultur (jen poznamenejme, že toto slovní spojení je přinejmenším mírně problematické, a to především ve vztahu k tzv. primitivním kulturám, jejichž jazyk je naprosto odpovídající jejich vlastním potřebám; rozhodně proto není primitivní) postupem času vyvinula své písmo na základě odrazu hlásek mluveného jazyka. Takové písmo pak kultuře sloužilo a slouží jako způsob uchování textů, který je jednoduchý a snadno osvojitelný pro většinu členů jazykového společenství. V kul-

turách, které hláskové písmo z tradičních důvodů nezavedly, se musí běžný uživatel výuce grafické soustavy intenzivněji věnovat. Používání nehláskových písem ale vysvětluje jeden jazykový aspekt. Např. u čínštiny, která disponuje nehláskovým psacím systémem, se na hláskové rovině její struktura vyznačuje oslabenou kombinatorikou hlásek. Čínština tvoří menší množství typů slabik, než je tomu u většiny dalších jazyků. To je vhodné pro užívání nehláskového písma, jazyková intuice u něj sleduje slabičný systém, ve kterém nepotřebuje klesat v popisu níž, k hláskám, z nichž se tyto slabiky skládají. Přístup mluvčích např. evropských jazyků k čínskému psacímu systému jako k nezměrně komplikovanému přímo odráží onu jazykovou intuici, která čínské mluvčí váže ke slabice a ty evropské k hlásce. Vzhledem k rozdílu mezi jazykovým systémem čínštiny a evropských jazyků se ale rozdíl v písmu obou kultur jeví jako přirozený.

Všechny dosud zmíněné vlastnosti hlásek se týkaly jazykové kultury, její historie a individuálního vývoje komunikačních dovedností člověka. Popsali jsme vztah jedné jazykové jednotky a společenství užívající jazyk, jehož je tato jednotka součástí. Lingvistika jako vědecká disciplína na hlásky nahlíží širěji, z hlediska jejich systémové úlohy v jazyce. Na tomto úkolu také ve fonologickém popisu vystavěla svou metodologii a díky tomu následně také vědní autonomii. Lingvistika se zajímá především o ty vlastnosti jazykových jednotek, které jsou odpovědné za fungování jazykového systému a odlišují jednu jazykovou jednotku od druhé. Podívejme se proto na hlásky i z tohoto systémového hlediska.

Lingvistický popis hlásky

Všechny přirozené jazyky se vyvinuly na základě rozdělení vlastního zvukového repertoáru na několik desítek hlásek. Množství hlásek jazyka se při jeho rozvoji pravděpodobně zvětšovalo, ale počet nikdy nepřekročil několik desítek. Tento jev lze snadno vysvětlit sémioticky. Jedná se o nutnou systémovou a ekonomickou podmínku existence jazykového systému, který má uživateli sloužit k orientaci ve vnějším

prostředí a ke komunikaci. Právě tento aspekt stavby jazykových systémů budeme dále rozvádět. Budeme muset ale nejdříve definovat základní sémiotickou jednotku, tzv. jazykový znak.

Znak je funkcí, můžeme říci asociovaným spojením, výrazu a obsahu – pro začátek můžeme říct, že se jedná o spojení významu znaku a toho, čím je znak vyjádřen, např. tvaru slova nebo značky. Pojmům výraz a obsah se budeme podrobně věnovat později. Musíme ale poznamenat, že tato definice znaku je dvousložková a představuje tzv. dyadický model znaku. Dyadický pohled na znak není jediný možný a dokonce můžeme tvrdit, že větší pozornost byla v historii vědy o znacích věnována modelu tvořenému třemi složkami (tj. triadickému), který s výrazem (značícím vehikulem, tzv. representamen) spojuje jednak interpretans (obsah znaku v užívání jedincem) a objekt. Objektem se zde nemíní fyzická entita, ale reference k tomuto objektu jakožto jednotlivé fyzické entitě i třídě těchto entit. Dyadický model, který budeme užívat, neřeší pragmatizování obsahu, tedy to, že je obsah aktuálně užíván subjektem sémiózy. Je proto výkladově úspornější a v konceptu obsahu v jistém smyslu zahrnuje objekt i interpretans; odhlíží naopak od kontextu užívání znaku.

Při popisu znakového systému vychází sémiotika z konceptu tzv. purportu či umweltu (ten zavedl L. Hjelmslev, 1972, respektive J. von Uexküll, 2006). Konceptům purport a umwelt rozumíme následujícím způsobem: každý uživatel sémiotického systému neboli subjekt sémiózy má určité percepční dispozice a vnímanou širší vnějšího světa. Vidí jen jisté spektrum barev, slyší jen v jistém frekvenčním zvukovém rozsahu atd. Tento recepční rozsah navíc subjekt sémiózy dělí do různých velkých částí, které je schopen rozlišovat, např. právě jako různé barvy či zvuky. Toto škálování percepčního rozsahu preartikuluje/předtváří to, s čím může subjekt sémiózy v sémióze (tedy při užívání znaků) nahlížet. Subjekt sémiózy může svůj purport/umwelt rozšiřovat např. za pomoci pozorovacích zařízení, dalekohledů, mikroskopů apod. Jimi si zvětšuje rozsah vnímaného světa a může jeho části nechat vstupovat do sémiózy jako znakové a případně komunikované obsahy. Subatomární jevy nebo galaktické formace jsou toho příkladem – mluvíme o nich na

základě rozšíření svých pozorovacích schopností. Purport/umwelt se liší u jednotlivých živočišných druhů – na to právě upozorňuje von Uexküll (2006). Vnímání světa lidmi, psy, netopýry, včelami a bakteriemi má rozličnou kvalitu, která se stává materií pro obsahy odlišované tím kterým živočišným druhem (von Uexküll přirovnává biosféru k symfonii, v níž jsou v souhře umwelty jednotlivých živočišných druhů a jejich obsahy). Zaměříme-li se na člověka jako subjekt sémiózy a jeho přirozený percepční rámec, je obsahovou sadou to, co ve vnějším světě vnímá a jak jednotlivé kvality odlišuje. Segmentace světa na jednotlivé typizované jevy slouží subjektu sémiózy k orientaci v prostředí. Tradičním příkladem je rozlišování mezi různými podobami sněhu Laponci. Ti rozeznávají a dokáží pojmenovat více druhů sněhů než obyvatelé jižnějších oblastí a na základě toho se pak řídí při svém jednání (tento příklad je většinou vzpomínán u problému tzv. jazykového relativismu, tedy u otázky determinace percepce a konceptualizace světa sémiotickým systémem subjektu sémiózy, kterou není radno přeceňovat ani podceňovat, ale v našem výkladu nemá rozhodující roli, tu hraje purport/umwelt sám o sobě; přeceňovat není radno ani tradičně udávaný počet zmíněných druhů sněhu). V této souvislosti uvádí Hjelmsov příklad s rozlišováním druhů lesa (Hjelmsov, 1972, s. 60). Kultura, která bude založena na životě v lesním prostředí, bude rozlišovat více druhů uspořádání stromů v krajině, než jen les, hvozd, remízek apod., a to samozřejmě proto, že je pro tuto kulturu takové rozlišení nutné nebo výhodné. V jiném příkladu strukturace obsahové sady Hjelmsov ukazuje (s. 59), jak dva odlišné jazyky rozdělují stejnou část barevného spektra. První jazyk spektrum rozděluje mezi čtveřici barev (zelenou, modrou, šedou a hnědou), druhý jazyk dělí stejné spektrum jen na barvy tři. V případě segmentace spektra barev se typicky jedná o nerozlišování zelené a modré barvy v jazycích některých severských národů, a to opět v souvislosti s purportem/umweltem (v severních zemích dopadá sluneční světlo vzhledem k povrchu planety pod větším úhlem, v důsledku toho člověk hůře rozlišuje barvy v rozsahu zelené a modré). Kvality vnějšího světa jsou tedy v jednotlivých jazycích různě tříděny v obsahových veličinách. Tyto obsahové veličiny pak napomáhají sub-

jektu sémiózy reagovat na stavy vnějšího prostředí, a to pro něj adekvátním a zároveň ekonomickým způsobem.

Obsahová sada představuje pro subjekt sémiózy určitou výhodu, protože redukuje složitost vnějšího světa na části, mezi nimiž musí rozlišovat a na něž různě reaguje. Znakový systém a sémióza se tak jeví jako nástroj pro zachování subjektu sémiózy, pokud si jej představíme jako fyzickou entitu ve vnějším prostředí. Znakový systém a sémióza je tedy nástroj pro záhodné jednání a usnadnění výhodné interakce s „nekonečně složitým“ vnějším světem. Obdobně jako obsahová sada je ekonomizovaná i sada výrazová, míra její ekonomizace je ale značně rozvinutější. Vysvětlení, proč tomu tak je, je možné na základě následujícího příkladu, ze kterého také vyplyne, proč přirozené jazyky vždy disponují hláskami.

Představme si dva filozofy, kteří se dohodli používat k dorozumívání mezi sebou jazyk neobsahující žádné hlásky. Jejich komunikace je velice blízká komunikaci na ostrově Balnibarbi z Gulliverových cest. Filozofové si místo toho, aby pojmenovávali předměty, tyto předměty ukazují. Tuto představu můžeme dále rozvést tak, že pro každý typizovaný předmět používají při komunikaci jeho předem připravenou maketu. Postupně vytvářejí makety všech druhů ovoce, živočichů, nástrojů, lidských povolání, společenských rolí, pocitů a dalších abstraktních pojmů. Mají-li s těmito maketami filozofové zacházet v aktuální komunikaci, musí si je na vhodném místě rozestavět kolem sebe a pro vyjádření obsahů je vždy ostentativně uchopit. Makety, které jsou příliš daleko, musí vždy nejprve donést, a až pak s nimi mohou vést komunikaci. Tento způsob komunikace značně znesnadňuje domluvu mezi oběma mluvčími. Obsahů, které chtějí filozofové v rozhovoru komunikovat, postupem času narůstá, až se vzdálenost, kam je nutné chodit pro makety, stává fyzicky náročnou. Pokud filozofové makety nevracejí na svá původní místa, tvoří z nich nakonec neuspořádanou zmeň, v níž se nesnadno orientuje. Maket jsou statisíce či miliony a akt jednoduchého sdělení kombinující plavání, želvu, harpunu, jídlo a útek představuje pro filozofa náročnou cestu. Filozofové nakonec použité makety na místě určeném ke vzájemné komunikaci pokládají vedle

sebe, ty se na sebe vrší, navzájem se odírají a rozbíjejí. Zničené makety je nutné znovu vyrábět. Alternativou by bylo opatrně makety znovu a znovu nosit na původní místa v nějak uspořádaném systému a riskovat tak stejně jejich případnou ztrátu. Ve chvíli, kdy filozofové musí komunikovat abstraktní pojmy, jako je krása, zisk, hlad, právo nebo spravedlnost, se stává dosažení takového pojmu hrou či kolekcí řad maket, které mají metaforicky či metonymicky vyjádřit zkušenost, jež nemá ve vnějším prostředí jednoduchou objektovou podobu či schéma.

Pokud mají filozofové na ostrově Balnibarbi žít, budou se muset vrátit k přirozenému jazyku, za pomoci kterého si smluvili svůj experiment, jinak dojdou k záhubě, například prostým vyhladověním či fyzickou likvidací od jimi navršených a z vysokých hromad padajících maket – to je samozřejmě hyperbola. Makety filozofům pro život na ostrově jednoduše nemohou stačit, z čistě pragmatického hlediska.

Myšlenkový experiment s filozofy ukazuje, že strukturace výrazové sady je také ekonomická a že taková být musí, aby se subjekt sémiózy shledával se svým znakovým systémem jako s užitečným nástrojem a s nutnou podmínkou přežití ve vnějším světě. Výrazy musí být subjektu sémiózy bezprostředně dostupné. Proto jsou konstruovány z toho, co subjektům nabízí jejich fyziologie. U člověka byla preferována zvuková podoba výrazu, která nezatěžuje jeho končetiny sloužící k pohybu a vykonávání nejrůznějších činností, nevyžaduje dokonce světlo a ani přímý optický kontakt. Volbu zvukové podoby výrazu lze proto vnímat jako velice výhodnou. Můžeme ji prohlásit za adaptaci, evoluční výhodu posilující zdatnost druhu (tedy člověka) v soutěži s jinými druhy v podmínkách prostředí, ve kterém žije.

Z tohoto pohledu by u filozofů na ostrově mohly mít obsahy znakovou funkci k výrazům, které by představovaly unikátní zvuky. Požadavek na jejich unikátnost by pak ale vzhledem k počtu obsahů kladl značně náročný úkol na fyziologii subjektů sémiózy – tedy na jejich artikulační orgány. Těžko si představíme nejen artikulační, ale i percepční ústrojí, které by se s takovou strukturou výrazů dokázalo vypořádat, myslíme-li samozřejmě na přirozeně utvořená fyziologická ústrojí. V přirozených podmínkách sémiózy z tohoto důvodu dochází k další

nevyhnutelné ekonomizaci výrazu. Tvořeno je jen malé množství značně odlišných zvuků, které dostatečně vysoké množství výrazů vytvářejí svými kombinacemi. V návaznosti na to můžeme následujícím způsobem formulovat jazykový princip arbitrárnosti: k obsahu je přiřazen libovolný výraz v podobě kombinace hlásek.

Arbitrárností výrazu přiděleného k obsahu se v jazyce vyhýbáme vytváření výrazů motivovaných podobností k jejich obsahům – maketám. Používáme výrazně odlišné zvuky, které nazýváme hláskami, z nichž se kombinují všechna slova, věty a texty našeho jazyka. Hlásky nemají, ačkoliv se to na první pohled nemusí zdát, žádný vztah k obsahu, který je jim ve znaku přidělován, jsou jen ekonomickým repertoárem prostředků, z nichž se znaky tvoří. Někdy můžeme najít podobnost mezi kombinací hlásek ve slově a obsahem – např. u onomatopoických slov. Hlásku /f/ nacházíme ve slovech jako foukat nebo flétna, která vznikla nápodobou vanoucího vzduchu, což u hlásky /f/ výrazně pociťujeme. Arbitrární povaha hlásek tak ale popřena není, a to vzhledem k tomu, že k nápodobě obsahu výrazem používáme v onomatopoických slovech vždy jen repertoár hlásek daného jazyka, žádné jiné zvuky. K nápodobě a maketám se tak sice přibližujeme, ale hlásky při tom neztrácejí svoji nejdůležitější vlastnost, a to je kombinace vybraných zvuků a z ní vycházející podoba výrazu (viz Faltýnek 2010, s. 80-83).

Požadavek principu jazykové arbitrárnosti naplňuje tzv. dvojí artikulace jazyka. Arbitrární jednotky (zvuky výrazu, tedy hlásky) k ničemu nerefekují a nemají tak relaci k obsahu, tyto jednotky se nazývají jednotkami druhé artikulace. Hlásky se dále kombinují do jednotek, které tvoří znak, tj. obsahovou a výrazovou relaci, tedy do jednotek první artikulace. Dvojí ustavení jednotek výrazu, jeho dvojí artikulace na hlásky a z nich kombinovaná slova (či morfémy), je sémiotickou nutností a logickým důsledkem systematizace jazyka pro jeho použitelnost.

Vlastnost přirozených jazyků disponovat hláskami popisuje jazykověda právě konceptem dvojí artikulace. Dodejme, že dvojí artikulace se vztahuje k rovině výrazu. Přenos konceptu dvojí artikulace do obsa-

hové roviny vytvářel sice přehledné, ale značně problematické způsoby popisu obsahu, historicky neuspěl a neodpovídá příliš tomu, jak jako subjekty sémiozy s obsahem nakládáme. Aby bylo možné vysvětlit, proč je výrazová rovina jazykového systému organizována na principu dvojí artikulace, bude nutné krátké představení dalších základních sémiotických východisek.

Ekonomizace výrazové sady se nezastavuje u repertoáru několika desítek hlásek. Kombinované jednotky výrazu, tedy hlásky, by sice mohly být zcela odlišné zvuky (písknutí, mlasknutí, syčení atd.), jak jsme psali výše, a jejich několik desítek by stále mohlo tvořit dostatečně ekonomický repertoár pro tvorbu všech výrazů. Hlásky ale zcela odlišnými zvuky nejsou, což je důsledkem usnadnění práce artikulačních a percepčních orgánů. Artikulační ústrojí moduluje zvuk a sleduje při tom několik kvalit, které se k modulaci nabízejí a které jsou zároveň těmi nejjednoduššími či základními možnostmi odlišení zvuků. Zvukové kvality výrazu jsou dány následujícími odlišnostmi (které však pro ilustrativnost zjednodušíme): výška šumu, rozsah šumu, změna šumové charakteristiky a její časový průběh, tónový průběh o určité frekvenci a jeho změna. K těmto zvukovým vlastnostem hlásek můžeme doplnit způsoby jejich vytváření: zvuk může být tvořen v artikulačních orgánech na různých místech vytvořením překážky proudu vzduchu a jejím uvolněním, ustavením určité úžiny, kterou vzduch prochází, nebo např. krátkým a nečastým kmitáním překážky v proudu vzduchu. Takto vytvořený zvuk nazýváme jako šum. V artikulačních orgánech se dále tvoří tón, který mohou artikulační orgány měnit z hlediska jeho výšky a rezonance. Podstatné vlastnosti odlišující hlásky mezi sebou se nazývají jako distinktivní rysy.

Jestliže se v každém jazyce vyskytuje několik desítek hlásek, pak popis odlišností, které jsou odpovědné za jejich jedinečnou podobu, obsahuje často méně než deset kategorií, které se na tvorbě hlásek podílejí. Tyto kvality jsou často postaveny proti sobě jako binární volby. Přítomnost či nepřítomnost jedné konkrétní zvukové kvality odlišuje jednu hlásku od druhé – např. kmitání hlasivek a tím způsobené zaznění tónu odlišuje hlásky /p/ a /b/. Tím se ekonomizace výrazové

sady znovu vysoce zvyšuje a zjednodušení repertoáru tvoření výrazů tak dosahuje velice elegantní podoby. Ekonomizace, tentokrát artikulační a percepční, se opět projevuje jako určující faktor sémiózy.

Když se hlásky kombinují ve slově, jejich zvukové vlastnosti se mění. Hlásky kombinované ve slově se často proměňují v jinou hlásku, mohou se v sérii hlásek ztratit, nebo naopak objevit. Například na konci slova se v mnoha jazycích znělé hlásky změny na neznělé, ztrácejí distinktivní rys znělosti (kmitání hlasivek a tím tvořený základní tón – slovo provaz se vyslovuje jako [provas]). Jestliže se v některých jazycích znělá hláška vyskytne v bezprostřední blízkosti neznělé, musí se jedna z nich právě v této vlastnosti připodobnit ke druhé – buď znělost ztrácí, nebo nabývá (slovo vstup se vyslovuje jako [fstup], slovo shoda [schoda] jako [zhoda]). Tento proces se dá snadno vysvětlit výslovnostní ekonomii, která usnadňuje výslovnost v tom smyslu, že pokud jsou rozkmitány či naopak nerozkmitány hlasivky u první z hlásek, druhá bude mít stejnou vlastnost, aby nebylo nutné hlasivky v jejich aktivitě náhle zastavovat nebo aby je naopak nebylo nutné náhle rozkmitávat. Tlak na toto usnadnění je čistě fyziologický, nemá proto přímý vztah k jazykovému systému, který se mu může na rovině hlásek ekonomizačnímu tlaku podvolit. Hlásky se tak nejeví jako dané a neměnné kombinace distinktivních rysů. Začínají naopak tvořit kombinatorní množinu svých podob, které se místo od místa v zapojení do řetězce hlásek mění. Česká hláška [ch] ve slově bůh [bůch] se ve výskytech v tomto slově střídá s hláskou [h] např. ve tvaru [boha], s hláskou [ž] ve tvaru [boží] a také s hláskou [š] ve tvaru [bošský]. Hláška [ch] se proto nejeví jako ostře vymezená veličina, vždy zvukově identická, ale jako pole různě kombinovaných distinktivních rysů, které jsou v kombinacích hlásek proměnlivé. Hláška se tak ustavuje na základě jejího vztahu k hláskovému okolí a k pozici v komplexní výrazové struktuře, např. vzhledem k pozici ve slově. Míra této variability hlásek se mezi jazyky liší a některé jsou vůči ní značně odmítavé. Čínština, vrátíme-li se k ní, má například velice specifické možnosti kombinace hlásek a z toho také vyplývá jejich nízká proměnlivost. Při vývoji čínské slabiky byl absolutizován zákon otevřené slabiky (v jiných jazycích, např. v češtině,

působil také, ale ne se stejnou důsledností). V závislosti na tom se možnosti tvorby čínské slabiky omezily na tendenci kombinovat konsonantní začátek s vokálním koncem slabiky (jsou zde samozřejmě výjimky, které ale otevření slabiky stále dodržují, např. nosové rezonance hlásek umístěných na konci slabiky). Jednotlivé hlásky čínštiny se proto mohou objevovat jen na určitých místech ve slabice. Malé množství kombinatorních možností pak způsobuje, že se tato malá míra variability hlásek v jejich kombinatorice přenáší na morfologickou rovinu, kdy omezené množství morfémů, kterých je jen tolik, kolik je možných slabik, má vysokou polysémií a homonymii. Čínské slovo se pak ztotožňuje s těmito morfémy, lexikální rovina poklesá k rovině morfologické a potřeby vyjádření jsou řešeny kombinací morfémů/slov v kompozitních slovních jednotkách. Podobně nízká morfologická variantnost angličtiny způsobuje její pevný slovosled. Výrazové distinkce se zde naopak delegují o jazykovou úroveň výš.

Ukázek proměnlivosti hlásek je možné u konkrétních jazyků uvádět mnoho, výše zmíněné příklady ji jen ilustrují. Podstatné je, že se hlásky projevují jako kombinace distinktivních rysů, které se v řetězci hlásek mohou měnit. Povaha hlásky se tak nejeví jako soubor distinkcí, ale jako kombinatorní pole jejich realizace. To je zcela odlišný pohled na hlásky než jako na jedinečné entity distinktivně odlišné od všech ostatních entit systému. Systém se ukazuje komplexněji, se zabudovanými pravidly proměn entity v entitu jinou.

Rozšíření systému hlásek o jejich pravidelné proměny v řetězci hlásek je umožněno jejich specifickou funkcí – hlásky rozlišují význam, samy jej však nenesou. Je důležité na hlásky nahlížet jen jako na nositele distinkcí mající v řetězci informační hodnotu. Informace je kvalita, která se projevuje jen na výrazové sadě, s obsahem nepřichází do kontaktu. Spočívá v pravděpodobnosti kombinace hlásek v textu. Z tohoto úhlu pohledu nemají hlásky vztah k významu, jejich vlastností je tzv. kolikátost, neboli postavení v systému všech hlásek, případně v jejich výčtu (jsou několikáté v seznamu). A tento systém musí obsahovat i pravidelné proměny hlásek, jak jsme je popsali výš. Na to se u hlásek často zapomíná.

Zavedení kombinatorických pravidel proměny hlásek do systému hlásek vychází vstříc artikulační a percepční ekonomii – děje se jako usnadnění výslovnosti i jako prostředek snadnějšího přijetí zprávy. Vývoj jazyků je často svázán s těmito změnami, určitá hláska se na základě nich může trvale změnit na jinou. Artikulační ekonomie ale není nejdůležitější podmínkou komunikace v přirozeném jazyce, tou je dvojitá artikulace a systém hlásek. To, jak se kombinují distinktivní rysy v hláskách, je jen míra volnosti, na níž vytváří tlak informační/komunikační úspěšnost a artikulační snadnost (myšleno technicky, z hlediska přenosu mezi vysílačem a přijímačem, při jejich percepční citlivosti a intenzitě šumu prostředí; rozdíly zvuků musí být snadno rozlišitelné a snadno artikulovatelné, a to i z hlediska předpověditelnosti výskytu v řetězci). Tyto změny ale nikdy nemohou popřít základní jazykovou vlastnost – dvojitou artikulaci.

Analogie hlásky v genetickém kódu

Na hlásky jsme prozatím pohlíželi jen z hlediska přirozeného jazyka. V přirozeném jazyce umí tyto jednotky každý z nás intuitivně identifikovat. Ukázalo se, že jejich přítomnost ve struktuře jazyka je důsledkem systémových a ekonomizačních podmínek sémiozy. Hlásky jsou nezbytnou sémiotickou podmínkou a univerzálií, ne jen konceptem, který známe z našich jazyků a kterému hledáme v jiných znakových soustavách analogii. Naše intuice při popisu hlásek tak nalézá obecnou podmínku znakového chování.

Když o hláskách uvažujeme, obracíme se především na jazyky, které používáme a známe, tedy národní jazyky, jako je čeština, čínština atd., případně se můžeme zajímat o exotické jazyky jako je laponština či indiánské jazyky. To je důsledkem postavení národního jazyka jako přirozeného sémiotického systému. V jednom případě však nejširší diskurz vědy tento přirozený pohled na jazyk přenesl na jinou sféru jevů – na genetickou informaci a její užívání organismem. Slovo jazyk při tom bylo zaměněno za techničtější, ale naprosto synonymní slovo

kód a celý proud poznání tohoto kódu se připoutal k nám vlastnímu pohledu na jazyk, intuitivnímu i vědecky řízenému.

Musíme předpokládat, že rovinu hlásek, tedy druhou artikulaci, nalezneme jak u přirozeného jazyka, tak u tzv. genetického kódu. Tento předpoklad vychází z požadavku, aby byl sémiotický systém pro svého uživatele výhodným nástrojem. Analogicky k situaci filozofů z Balni-barbi, jejichž makety napodobující stavy vnějšího světa nejsou pro život na ostrově dostatečné. Sémiotický systém proto potřebuje arbitrární sadu výrazu dostupnou uživateli. Takovou sadu dovolují konstituovat hlásky, nic neznamenající a fyziologicky snadno produkovatelné jednotky, které v kombinaci tvoří vyšší výrazové jednotky. Až ty jsou v asociativním vztahu k významu, čímž utvářejí znak. Jak ale k analogii s hláskami ve vztahu k DNA a jejím funkcím v organismu přistoupit? Rovinu druhé artikulace sémiotický systém organismu obsahovat musí, jak ji ale správně popsat, když nás k ní bude vždy přivádět analogie s hláskami, s druhou artikulací našich přirozených jazyků?

Jazyková analýza vychází z nalézání výrazů jazyka a jim odpovídajícího obsahu, to znamená ze znakové funkce. Jedině změna znakové funkce, to znamená navázání jiného výrazu s jiným obsahem, může být kritériem pro identifikaci veličin/jednotek obsahu i výrazu. Pokud bychom měli metodu jazykovědy vyložit ve zkratce, ale pořád dostatečně adekvátně, mohli bychom formulovat toto:

Jazykověda vychází z textu, který představuje materiálovou danost. Na základě té musí jazykověda dokázat, jakými částmi a vztahy je text tvořen. Má-li jazykovědec text, postupuje za tímto účelem jednoduchou procedurou. Text dělí na jednotlivé části. Tyto části v daném místě výskytu v textu střídá se všemi ostatními možnými částmi. Pokud tímto střídáním docílí změny znakové funkce, identifikuje zároveň jednotky. To znamená, že pokud v textu provede dělení, ale nenajde v substituci dvou oddělených částí novou znakovou funkci, musí se v proceduře vrátit a dělení nepovažuje za platné. V postupném dělení textu dochází jazykovědec ke stále omezenějšímu množství jednotek, které může na místě zkoumané jednotky střídát, a to tak dlouho než dosáhne poslední úrovně. Poslední úroveň je velice omezená a skládá se

z jednotek, které význam samy o sobě nemají – to jsou hlásky. Postup procedury, kterou jsme zde popsali, je uplatněním tzv. komutačního testu na všech rovinách (od fonologické až po syntaktickou či textovou) a představuje jistou hyperbolizaci čistě jazykovědného postupu při analýze textu (lze tak dobře popsat fonologický systém, na úrovni věty je ale tento přístup značně neoperabilní). Tento postup není možné ve všech důsledcích uplatňovat, metodologicky ale představuje čistý koncept, který bezproblémově vystihuje systém jazyka stojící za textem.

Chceme-li postupovat naznačeným jazykovědným způsobem tak, abychom se dobrali úrovně hlásky v sémióze organismu, musíme se nejdříve pokusit o popis obsahové a výrazové roviny této sémiózy. Je třeba vysvětlit, proč je možné tyto roviny považovat za provázané ve znakové funkci a co tuto provázanost zajišťuje. Ve všech krocích ale bude nutné následovat ekonomizační principy, které se v předchozím výkladu o hláskách ukázaly jako nutné podmínky existence sémiózy a jako vysvětlující momenty v hledání jejího systémového uspořádání.

Obsah sémiózy organismu lze definovat, ale se značnými problémy při výběru jeho možných popisů. Veškeré chování organismu vychází z reakce na jeho vnější a vnitřní prostředí. Vypsát veškeré fyzikální aktivity, které konkrétní jednání organismu představují, není možné. Sledovat každý atom a každou změnu vztahů mezi molekulami nepřichází v úvahu. Můžeme to ale nahradit popisem funkce – celkovou reakcí organismu na určité fyzikální stimuly (behaviorálním popisem), případně cílem orientovaného jednání (teleologickým popisem). Pokud například bakterie postupuje v gradientu glukózy, nemusíme popisovat veškeré s tím spojené fyzikální aktivity, reakční dráhy atd., ale vyjádříme toto chování funkcí. Můžeme se uchýlit k popisu cíle orientovaného jednání, který celou síť fyzikálních aktivit organismu zahrne – v případě bakterie a glukózy je to získání potravy. Zdá se, že funkcionální výklad je nutnou podmínkou smysluplného popisu chování organismu. Rozdíly v chování organismu pak tvoří obsahovou sadu, kterou lze definovat jako organismem odlišované kvality v jeho purportu/umweltu, k nimž v popisu zahrnujeme reakce organismu na ně. Bakterie například nemusí vůbec rozlišovat své reakce na zvýšený výskyt se-

lenu a olova v prostředí. Na oba tyto stavy světa bude reagovat naprosto stejným způsobem, pohybem pryč. Vrátime-li se k příkladu s barvami, typy lesa a typy sněhu, kterými je výše vysvětlena organizace obsahové sady, je bakterie a zmíněné kovy vhodný příklad pro purport/umwelt organismu. Jako Švédové nerozlišují modrou a zelenou, je pro bakterii selen a olovo jednou veličinou světa, na kterou reaguje, není mezi nimi schopna rozlišovat.

U příkladu s bakterií a glukózou by se mohlo zdát, že je nutné dále a důmyslněji vysvětlit, co je obsahem v sémióze buňky. Podstata obsahové sady zde ale zůstává stejná jako u filozofů na Barnibalbi. I v tomto případě se musíme snažit obsahy vystihnout vhodným metajazykem, i když při výkladu chemických procesů v organismu to nemusí být samozřejmé. Subjekt sémiózy má ve svém purportu/umweltu vyděleny části světa, na něž reaguje jako na odlišné stavy, pro sebe výhodným způsobem. Změna perspektivy zaměřením se na bakterii s jejím purportem/umweltem nemohla nic změnit.

K chování organismu, k různým funkcím, které naplňuje, se na fyzikální rovině přímo vztahují proteiny. Tvar proteinu z něj činí nástroj vhodný ke konkrétní činnosti – transportní, receptivní, mediující interakci dvou přímo neinteragujících chemických sloučenin, přemostující určité fyzikální bariéry, stavební atd. Právě u proteinů se proto setkává popis funkce a její fyzikální podstata v podobě tvaru. Vztah tvaru a jeho funkce nám dovoluje hledat vhodný metajazyk pro popis struktury obsahu. Na rovině buňky se díky tomu můžeme pokusit postoupit dál směrem k adekvátnímu uchopení obsahu. Proteiny jsou tzv. peptidy, řetězce aminokyselin, které v prostředí zaujmají určitý tvar, tzv. konformaci. Představíme-li si je jako trojrozměrné objekty, bude za jejich funkci zodpovědná některá jejich část či více částí. Jeden specifický tvar se např. může vázat s konkrétní sloučeninou a činí tak z proteinu specializovaný nástroj pro výkon určité funkce v organismu. Případně se tyto funkce v proteinu sdružují – jedna část tvaru proteinu koná jednu funkci, druhá další a jejich společný účinek je funkcí celého proteinu.

U problému určení obsahové sady se tak dostáváme na nejnižší možnou úroveň. Funkci organismu v reakci na vnější prostředí zastupuje

konkrétní protein a jeho funkční části (samozřejmě v součinnosti s celou sítí dalších fyzikálních reakcí). Zaměříme se nyní na to, jak se protein s jeho funkčními tvary objevuje v organismu. Řetězec aminokyselin, z nichž se protein skládá a který se za mnoha podmínek konformuje do funkčního celku, vzniká translací řetězce mRNA. Ta je po řadě posttranskripčních procesů proměněným zápisem z DNA, který je využit jakožto formule k vytvoření aminokyselinového řetězce. Řetězec aminokyselin bezprostředně nabývá prostorovou podobu, přestává mít funkci jako sekvencionální sled jednotek a stává se prostorovou entitou s funkčním potenciálem interagovat s danými objekty. Popis výstavby proteinu na základě zápisu v DNA ukazujeme v instrumentáriu molekulární biologie v následující kapitole.

Pro výstavbu proteinů je v organismu historicky ustaven molekulární objekt, který na jedné straně váže řetězec mRNA a na druhé váže aminokyseliny. Tímto objektem je tRNA. Podobně jako protein může tRNA ve vzájemné interakci spojovat fyzikální objekty, které by se jinak než takto řízeným způsobem nesetkaly. Snad není předčasné tvrdit, že podobně jako u hlásek je na tomto místě možné položit arbitrární hranici obsahu a výrazu – tRNA k sobě vztahuje nemotivované (fyzikálně samostatně neinteragující) řetězce aminokyselin proteinu a řetězce RNA (respektive DNA, z níž je RNA transkribována). Když jsme u funkčních míst proteinu dosáhli nejnižší možné strukturační obsahové sady, musíme být už teď, když jsme identifikovali arbitrární přiřazení obsahu a výrazu, blízko hláskám sémiózy organismu – tak to alespoň napovídá ekonomický a systémový status sémiózy, který se v průběhu výkladu ozřejmil.

Funkce proteinu je dána konformovaným řetězcem aminokyselin. V tomto řetězci je těžké uvažovat o roli jediné aminokyseliny. Funkci proteinu můžeme destruovat tím, že v procesu tvorby proteinu zaměníme jednu aminokyselinu. Její záměna ale není bezprostředním původcem ztráty funkce. Takovou ztrátu způsobuje celý řetězec, v němž konkrétní aminokyselina porušila své vazby k sousedním aminokyselinám. Dopad této vazby se projeví v tvaru celého řetězce. Jedna aminokyselina tak představuje funkční prostředek, který má v kombinaci

s dalšími aminokyselinami jistý význam pro chování proteinu v organismu. Snad se náš výklad nepřiblížil k hláskám příliš chvatně – v tomto gardu však sémiotický výklad proteosyntézy vypadá korektně a zdá se, že identifikuje hlásky.

Jediná aminokyselina nemá žádný parciální význam v celkovém významu či funkci proteinu. Nemůžeme jí připsat roli určitého vztahu k vnějšímu světu, samostatně nezajišťuje reakci organismu na vnější prostředí. To jí přísluší až v určitém zapojení do řetězce. Tento test je analogický k popsané jazykovědné metodě – dělíme řetězec na části, zaměřujeme se na nejmenší možný úsek, jemuž můžeme přisoudit obsah. Jestliže tento úsek znovu rozdělíme na části, které při záměně změní obsah celého úseku a samy nemají obsah, nalézáme hlásky.

Pokud budeme na základě zkušeností se strukturou přirozeného jazyka chtít analogizovat s hláskovou rovinou i vyšší roviny, musíme konstatovat několik výchozích předpokladů. Bezprostředně vyšší úroveň může být jednak morfém, což je u přirozených jazyků nejtýpější. Příklad čínštiny ale ukázal, že morfémy a slova mohou mít homonymní, stejnotvarou podobu, že se tyto jazykové roviny mohou prolnout. Sémioticky není naprosto nutné, aby se mezi rovinou výpovědi a hlásek instalovaly v popisu dvě další úrovně, nic v systémovém ustavení sémiózy to nepředpokládá. Morfologická úroveň může být zcela zanedbána a za funkci slov pak odpovídají syntaktická pravidla, nerozhoduje o ní tvar slov.

Další analogie s jazykovými jednotkami dál rozvíjet nebudeme, bylo ale nutné, abychom naznačili, proč se u aminokyselin v řetězci proteinu shledáváme s druhou artikulací, nejnižší jazykovou úrovní kombinovaných nic neznamenajících jednotek, s hláskami. Vyšší jednotky, včetně vět a textů, musí být výslednicí jejich kombinace, i když v rozsáhlých a funkčně vzájemně složitě propojených komplexech (ale tak je to u textů přirozeného jazyka také), a tyto komplexy mají intencionální povahu, povahu chování, o něž nám od počátku jde.

O sémantice proteinového řetězce uvažovali např. Trifonov a Berezovsky (2002), Berezovsky – Kirzhner – Kirzhner – Rosenfeld – Trifonov (2003). V jejich přístupu je patrný náznak konstrukce metajazyka,

který by popsal menší části aminokyselinového řetězce a jeho funkce v prostředí buňky. Analytický postup těchto textů a řešení popisu struktury obsahové sady podírají výklad v přechozím odstavci. Potvrzuje se tím, že nejmenší částí organismu buňky, která má obsahovou podstatu, to znamená, že organismu umožňuje reakci na vnější prostředí, je nějak minimálně dlouhý řetězec aminokyselin.

V předchozím výkladu jsme označili tRNA jako adaptor neboli molekulární objekt umožňující interakci mRNA s aminokyselinami a zajišťující na základě řetězce bází výstavbu řetězce aminokyselin. Toto zprostředkování vztahu mRNA a aminokyselin představuje analogii arbitrárního přidělení obsahu a výrazu v přirozených jazycích. Jako obsah sémiózy organismu jsme charakterizovali funkce, které organismus provádí a které mu umožňují záhodné chování. Obrátíme-li se nyní k řetězcům DNA, můžeme je označit za výraz, který je vztažený k obsahům právě přes hranici onoho adaptoru. Slova a věty přirozeného jazyka jsou arbitrárně vystavěné z hlediska relace obsahu a výrazu, jsou složeny z hlásek. Všechny tyto aspekty sémiózy jsou na buněčné úrovni stejně tak přítomné. Organismus disponuje zápisem v DNA, který různě užívá a užívá jej proto, aby se rychle a snadno dostal k procedurám, kterými naplní své intence a díky kterým může náležitě interagovat s vnějším prostředím. DNA představuje rezervoár možných výrazů, které vstupují do vztahů s obsahy, tj. s tím, co organismus ve světě, vlastně ve svém purportu rozlišuje jako rozdílné stavy. Výrazová rovina ale musí zahrnovat veškeré proměny řetězce DNA, včetně sestříhových faktorů, faktorů exprese genů ale i konformaci proteinů atd. Vše, co se uděje v organismu před překročením hranice adaptoru, je výraz. To je formulace, která je např. pro biosémiotické znakové modely organismu nová. To vnímáme vzhledem k současným znakovým popisům živého jako překvapení, když se stanovení výrazu a obsahu přes rovinu adaptoru jeví být jako jediný sémioticky smysluplný přístup.

U přirozeného jazyka je vztah obsahu a výrazu tzv. asociativně spojený. U sémiózy organismu jsme asociaci nahradili adaptorem, jehož arbitrární přiřazování je zcela analogické znakům přirozeného jazyka.

Řečeno jednoduše, k nutným interakcím fyzikálních veličin řídících se fyzikálními zákony dodává další pravidla interakce, dříve ustavená a stále udržovaná pro jistou záhodnost v celku fyzikálních interakcí organismu, který právě jako celek postavený proti zbytku fyzikálního světa usiluje o své zachování.

Užití znaků je okamžité. V komunikaci promlouváme a řadíme v čase znaky za sebe, výrazy se objevují aktuálně vztažené ke svým obsahům. U organismu by se mohlo jevit, že pokud můžeme studovat DNA samostatně, jako text, setrvává výraz mimo obsahy. Pokud ale odpovídáme z pozice subjektu sémiózy, tedy organismu, je znakem vždy využitá potence DNA pro vytvoření záhodného chování. DNA je jinak mrtvá hmota, která znakově čeká na své uvedení do funkce k obsahu, jímž je využití při chování subjektu sémiózy – i sémióza organismu je tedy aktuální, bezprostřední interakce obsahu a výrazu. Analýza řetězce bází v DNA je procedura, kterou čeká souuvztažení s obsahem, a to je úloha, která má příliš mnoho proměnných v řešení. Ale to je tak u znakových systémů vždy, právě díky obsahu, o kterém musíme spekulovat a výrazu, který hledáme v celkové odpovědnosti za obsah.

Na konci úvahy o hláskách můžeme konstatovat, že jako rovina hlásek se na úrovni DNA vyjevují triplety bází, tzv. kodony. To jsou ty kombinace, které v tzv. genetickém kódu určují konkrétní aminokyselinu. Báze jako součásti kodonů jsou jejich distinktivními rysy. V analogii k výkladu o hláskách z první části textu teď pokračujeme rychle, dovoluujeme si to proto, že výše se již jasně ukázalo, že je pro sémiózu nutné ekonomicky nastavit strukturu distinktivních rysů – je to opět sémioticky systémový předpoklad a nedá se v sémióze obejít.

Kombinace bází je určující pro navázání aminokyseliny v transkripci. Považujeme-li kombinace bází za distinktivními rysy a budeme-li se dál řídit analogií s přirozenými jazyky, můžeme říci, že se nějak podobají jejich vlastnosti? Směrem od roviny hlásek k distinktivním rysům klesá počet funkčních veličin. Hlásek je v přirozených jazycích několik desítek a distinktivních rysů cca deset – redukce je značná. Kombinací, které tvoří triplet, je 64. Báze obsažené v tripletech jsou

čtyři a tvoří dvanáct možných kombinací, jež považujeme za distinktivní rysy. Mnohé triplety mají stejnou funkci – analogicky k proměnám hlásek ve slově, které jsme na mnoha příkladech ukázali výš. Tento aspekt je v přirozených jazycích čistě produkční, je způsoben usnadňováním artikulace a v jazyce nemá žádnou systémovou funkci. U kombinací bází vztažených k aminokyselinám je to podobné. V evoluci bylo využito pouze dvacet aminokyselin, a všechny triplety byly těmto aminokyselinám přiděleny jako využitě varianty zajišťující stabilitu přenosu genetické informace k proteinům a k potomkům.

Výklad kombinace bází jako distinktivních rysů naznačuje další shodu s hláskovým systémem jazyka. Hlávky nejsou různé zvuky, ale zvuky o různé přítomnosti a nepřítomnosti distinktivních rysů. Pokud by báze byly zcela rozličné chemické sloučeniny, museli bychom si představit značně komplikovanou sadu výrazu. Různé chemické prostředky by musel adaptor rozpoznávat na základě dalšího velkého množství pravidel zajišťujících vysoce nepravděpodobnou fyzikální interakci. DNA je ale konstruována na základě rozdílů prvků, které mají jednoduchou skladebnost a mohou se vyskytovat v jediném řetězci. Stejně jako je nevýhodné, aby byly hlávky zcela rozdílné zvuky, případně tvary, objekty atd. (viz reprezentace hlásek písmeny, o tom ale dále), kombinuje výrazová rovina organismu veličiny příbuzné, které se liší změnou jedné své vlastnosti. Produkční zdůvodnění stojí sice mimo systémová pravidla sémiózy, je ale nutné pro vysvětlení vyslovitelnosti u hlásek mluveného jazyka, v organismu pro udržitelnost výrazu v buněčném prostředí jedné buňky.

Biosémiotický výklad znakového chování organismu se k jazykovým analogiím uchyluje často. Je to přirozené, protože naši přirozenou zkušenost se znaky přenášíme na jinou sémiózu. Výklad znaku v biosémiotice vycházel většinou z popisu DNA a její struktury. A jestliže se jako intuitivně nejbližší jazyková jednotka nabízí hlávka, postupovalo se i zde tímto způsobem. Nejmenší funkční část DNA se vždy jevila jako snadno zachytitelná, báze jsou její stavební kameny, a tak byly analogizovány s hláskami. To je ale z hlediska určení výrazové a obsahové roviny organismu, jak jsme jej popsali, fatální lapsus. A je to také

hlavní příčina toho, proč se všechny znakové i informační analýzy DNA a organismu setkaly v naprostém vzájemném nedorozumění. Roli hlásky na výrazové rovině hrají triplety/kodony, to je sémiotická nutnost, navíc potvrzená funkční analýzou obsahové sady a jí odpovídajících výrazů. Báze jsou jen, jako znělost a neznělost, rozlišující vlastností svých kombinací, které zajišťují smysluplné akce organismu. Snad se nám teď už bude se znaky v buňce pracovat lépe a radostněji.

Báze nejsou písmena

Genetický kód

Počínaje popsáním struktury DNA mluví biologie, nově vznikající molekulární biologie a obecně věda a společnost o genetickém kódu (viz k tomu Watson – Berry, 2003). Dnes je genetický kód všeobecně intuitivně potvrzovaným vědeckým poznatkem. S předpokladem existence genetického kódu dnes bezprostředně vnímáme živé bytosti a biosféru. Na tento kód se pohlíželo s jistými předpoklady: Byly formulovány některé jeho základní vlastnosti a postupem času se pevně ustavilo nahlížení na to, jaký design tento kód má. Výše jsme odkázali k publikaci ozřejmující okolnosti objevu struktury DNA, jejímž autorem je J. Watson. Vlastnosti DNA a genetického kódu ve svých publikacích a přednáškách představoval, vysvětloval a popularizoval Francis Crick. Využijeme jeho textů jako reprezentativních pro představení dnes běžného pohledu na genetický kód. Vlastnosti genetického kódu Crick formuluje následovně.

Genetický kód k sobě vztahuje aminokyseliny (z nichž se skládají proteiny) a báze, které obsahuje DNA (Crick 1962, s. 11-12; Crick 1968, s. 367). V procesu výstavby proteinu se k sobě vztahuje 20 aminokyselin a 64 trojkombinací bází, tzv. tripletů (Crick 1968, s. 368). Tento kód je univerzální a až na výjimky, které ale nevybočují z principu vztahu bází a aminokyselin, je společný všem živým organismům (Crick 1962, s. 8; Crick 1968, s. 369).

Báze jsou v genetickém kódu spojeny s aminokyselinami arbitrárním vztahem. Ze všech možných aminokyselin je použito jen konkrétních dvacet, které se vztahují k bázím na základě zprostředkování jistými molekulárními prostředky (tzv. adaptorem – tRNA; Crick 1967, s. 342–343), přičemž mezi aminokyselinami a bázemi není přímá chemická afinita, mluví se zde o zmrzlé náhodě (Crick 1968, s. 369). Genetický kód je redundantní, a to v tom smyslu, že více tripletů odpovídá jedné aminokyselině. Některé tripletety naopak vztah k aminokyse-

linám nemají a slouží pouze jako signály vymezující hranici začátku a konce využitelné genetické informace. Relevantní roli v tripletech hrají především první dvě báze. Třetí pozice v tripletu často umožňuje variovat báze, aniž by došlo k záměně aminokyseliny. Umístění bází v tripletu tedy není náhodné. Crick (1968, s. 369) tuto systematickosti detailně popisuje.

Zápis bází je lineární, čte se v jednom směru a bez možnosti přeskokování bází (Crick 1966, s. 332–333; Crick 1964, s. 9). Má pevný čtecí rámec, v němž zachovává hranice tripletu. Zároveň se nepřekrývá (z aj. *overlapping*), to znamená, že nenese více informací současně (např. Trifonov ale kód chápe jako překrývající se, důvody popisuje s dalšími spoluautory v práci Popov – Segal – Trifonov 1996, s. 66; viz k tomu především Trifonov 1988, s. 508–510). Proces výstavby proteinu, tzv. proteosyntéza, probíhá (informačně a energeticky) směrem od bází k proteinům. Tento princip je nazván centrální dogma.

Uvedený popis genetického kódu byl dále precizován objevy molekulární biologie. Proces proteosyntézy byl popsán s mnoha dalšími proměnami původního řetězce DNA (do hnRNA, mRNA, v souvislosti s interakcemi s snRNA atd.) a procesy směřujícími k finálním produktům proteosyntézy a jejich funkcím. Byly popsány procesy sestříhu, jejich variace mezi organismy a částmi organismů, konformační procesy proteinů, metylace řetězců DNA, chromatinové interakce atd.

Genetický kód je vnímán jako lineární zápis bází vztahující se k tvaru a funkci proteinu. Báze jsou v praxi charakterizovány jako písmena (Crick 1967, s. 331; Crick 1962, s. 16). Tato písmena tvoří tripletu (trojice bází), které jsou pojímány jako slova, tzv. kodony – kódová slova. Tato slova mají kódem zprostředkovaný vztah ke konkrétní aminokyselině. Soubor tripletů (kódových slov) tvoří gen, jednotku, která má vztah k celému proteinu, jeho tvaru a z něj plynoucí funkci v organismu (Crick 1962, s. 8; Crick 1964, s. 2). O bázích jakožto písmenech a tripletech jako slovech se hovoří např. v následujících publikacích (vybíráme reprezentativní doklady pouze pro ilustraci): Stanford (1975, s. 74) ve svých Základech biofyziky hovoří o tripletech jakožto třípísmenných slovech, Weaver (2002, s. 569) ve své Molekulární biologii

říká, že kodony jsou kódová slova a skládají se ze tří písmen, Twyman (1998, s. 205) ve své syntéze molekulární biologie označuje báze jako písmena, triplety jako slova a geny jako věty a stejně tak se vyjadřují i Hartl a Ruvolová (2013, s. 10).

Genetický kód a přirozený jazyk

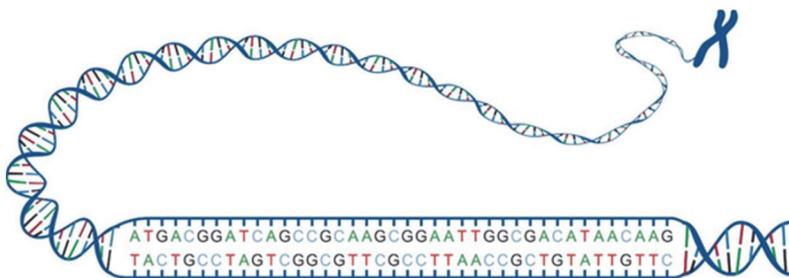
O genetickém kódu se vždy uvažovalo ve vztahu k přirozenému jazyku. Vědní obor molekulární biologie je pevně svázán s pojetím DNA a proteosyntézy jakožto analogie přirozeného jazyka. Analogie molekulárně genetických procesů s jazykem jsou v molekulární biologii do jisté míry instrumentální, představují tradiční přístup k terminologii. Částečně ale analogie s jazykem molekulární biologii zprostředkovává přístup ke genetickému kódu, jeho struktuře a funkci. Užívání jazykové metafory v molekulární biologii názorně ukázal Raible (2001). Ten provedl korpusové šetření na desítkách tisíc molekulárně genetických textů, z něhož jasně vyplývá, že termíny analogické k popisu přirozeného jazyka (písmeno, slovo, čtení, zápis, překlad atd.) jsou široce používány v běžné praxi této vědy. Searls (2002) ukázal, že molekulárně genetický výzkum nesdílí s lingvistikou jen terminologii, ale i výzkumné metody. Jakobson (1971, s. 655–696) dokonce potvrdil molekulární biologii korektnost využívání analogie genetického kódu a jazyka a hovořil přímo o struktuře genetického kódu, který se dle něj skládá z písmen (bází), slov (tripletů) a vět (genů). Dále poukazuje na to, že v genetickém kódu nacházíme vlastnosti jako je synonymie, suprasegmentální nebo syntaktická delimitace, systém distinktivních rysů či pružná stabilita. Jakobsonův jazykový výklad genetického kódu je pak přejímán dál (např. Katz 2008). O jazykové metafoře v biologii referuje Markoš a Faltýnek (2011).

V tomto textu chceme ukázat, že běžně přijímaný design genetického kódu, jak jsme jej představili výše, je možné zpochybnit. Domníváme se, že pojetí struktury genetického kódu, a to především v jazykové analogii, je od prvopočátku chybné. Chceme jej odmítnout, a to na základě popření analogieází a písmen. Ze sémiotického hledis-

ka jsme to již udělali (viz Faltýnek 2012). K tomuto účelu využijeme metodu kvantitativní analýzy textu, kterou představíme níže. Nejdříve ale čtenáře seznámíme se základním instrumentáři molekulární biologie, které je nutné k vyložení našich závěrů.

Instrumentárium molekulární biologie

Dědičná informace, obsahující instrukce k výstavbě organismu a řízení jeho biologických pochodů, je fyzicky zapsána v každé buňce ve formě deoxyribonukleové kyseliny mající podobu dvoušroubovice a známé pod zkratkou DNA. Způsob, jakým DNA uchovává informace, je založen na principu variací čtyř specifických makromolekul nukleových kyselin, konkrétně thyminu (T), guaninu (G), adeninu (A) a cytosinu (C). Střídáním těchto tzv. bází dochází k záznamu informace obdobně, jako když Morseova abeceda zaznamenává informace střídáním teček a čárek. Každá z bází má svůj chemicky afinitní (vzájemná vazba vodíkovými vazbami) protějšek, thymin stojí v DNA vždy proti adeninu a guanin proti cytosinu.

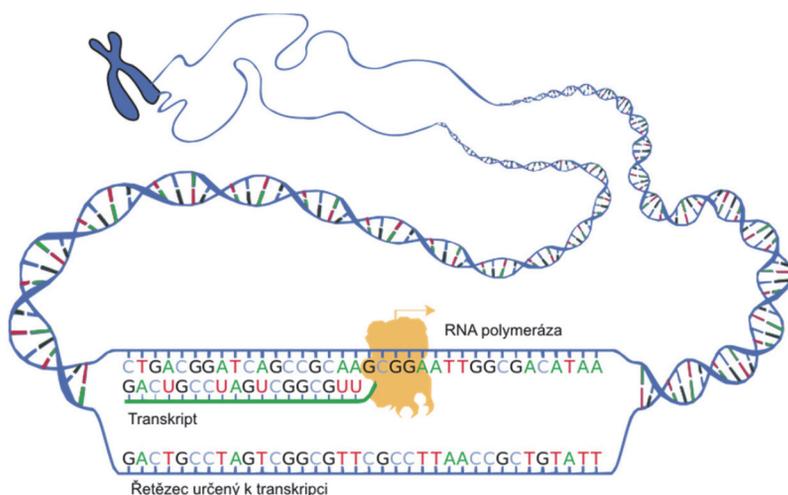


Obrázek 1: Dvoušroubovice DNA s jejími dvěma rameny nesoucí jednotlivé báze. A značí adenin, T thymin, G guanin, C cytosin. Dvoušroubovice DNA zaujímá strukturu v tzv. chromozomu (viz: <<http://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85259>>).

Lineární zápis bází DNA nám dovoluje celou DNA přečíst a přepsat formou textu, tj. zapsat zleva doprava, písmeno za písmenem (viz např. Cvrčková 2006, s. 17). Takový přepis je kopií řetězce bází zkoumané

DNA. Tento postup je v praxi nazýván jako sekvenování (detaily viz in Berg – Tymoczko – Stryer 2012, s. 140–148). Sekvenováním je řetězec DNA zprostředkován ke zkoumání nejrůznějšími nástroji.

Proces, kterým se z lineární sekvence DNA (tedy z určitého textu) stane protein – reálný fyzický nástroj využitelný v organismu – označujeme jako proteosyntézu. Celý tento proces můžeme popsat v několika krocích (viz např. Alberts et al. 2008, s. 329; Weaver 2002, s. 39):



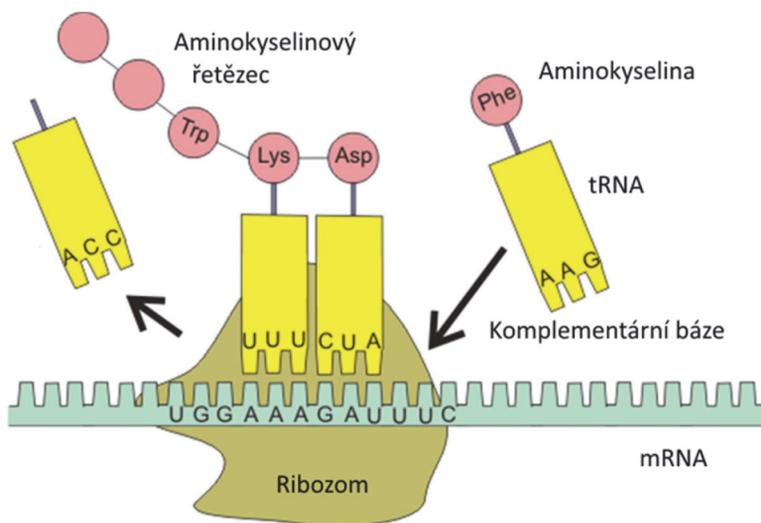
Obrázek 2: Transkripce. Přepis sekvence DNA na komplementární protějšky jejich bází RNA polymerázou, thymín je přepisován na uracil, zbylé báze na své komplementární protějšky (viz: <http://commons.wikimedia.org/wiki/File:DNA_transcription.svg>).

1. Transkripce. Po naplnění specifických podmínek uvnitř buňky se na začátek sekvence DNA (tzv. genu) připne protein RNA-polymeráza, který se po této sekvenci pohybuje v zadaném směru a nukleotid po nukleotidu tuto sekvenci přepisuje. Výsledkem je tzv. transkript – samostatná „pracovní kopie“ DNA ve formě ribonukleové kyseliny RNA, která je určena k zamýšlenému použití v proteosyntéze.

2. Úprava transkriptu. Transkript může být dále upraven (například vystříhnutím částí, které jsou v sekvencích vloženy a slouží jiným úče-

lům). Finální verze transkriptu, tzv. mRNA, je pak přemístěna k ribozomu, kde je konstruován protein.

3. Translace. Ribozom čte mRNA lineárně po trojicích nukleotidů (tripletech). Každému tripletu mRNA je donesena pro něj specifická aminokyselina, která je připojena k předchozí. Takto vytvořený aminokyselinový řetězec se při výstupu z ribozomu začne na základě fyzikálních vlastností jednotlivých konstituentů a fyzikálních vlastností molekul v prostředí (tedy na základě určitého kontextu) formovat, dokud nevytvoří stabilní konformaci (tvar) proteinu. Funkce a vlastnosti proteinu jsou determinovány fyzikálními vlastnostmi jeho makromolekulární konformace (Twyman 2004, s. 103).



Obrázek 3: Translace. Proces vzniku aminokyselinového řetězce proteinu. Tripletům mRNA je na ribozomu přiřazována tRNA s komplementárním tripletem, která nese aminokyselinu. Takto přiřazené aminokyseliny tvoří řetězec budoucího proteinu (viz: <Boumphreyfr/Wikipedia>).

Tripletům je aminokyselina přiřazena pomocí zprostředkujícího elementu – tzv. adaptorové molekuly tRNA. Adaptor tRNA získává svůj tvar při transkripci z DNA (Weaver 2002, s. 51). Funkcí této adaptorové molekuly je vázat na jedno ze svých vazebných míst specifický

triplet a na své druhé vazebné místo konkrétní aminokyselinu. Vztah aminokyselin a nukleotidů je tak zapsán přímo v DNA. Popsaný vztah byl nazván jako genetický kód (Weaver 2002, s. 12).

Genetický kód je tvořen variacemi čtyř různých nukleotidů v každé pozici tripletu. Triplet může nabývat 4^3 (64) možných unikátních kombinací, které tak mohou kódovat 64 různých aminokyselin. Všemi šedesáti čtyřmi realizovanými triplety je kódováno dvacet různých aminokyselin (Alberts et al. 2008, s. 367), mnoho tripletů kóduje stejnou aminokyselinu (kód je tzv. degenerovaný, viz obrázek 4). Některé triplety mají využití jako označení počátku a konce kódující sekvence.

1. pozice	2. pozice				3. pozice
U	C	A	G		
U	Phe Phe Leu Leu	Ser Ser Ser Ser	Tyr Tyr stop stop	Cys Cys stop Trp	U C A G
C	Leu Leu Leu Leu	Pro Pro Pro Pro	His His Gln Gln	Arg Arg Arg Arg	U C A G
A	Ile Ile Ile Met	Thr Thr Thr Thr	Asn Asn Lys Lys	Ser Ser Arg Arg	U C A G
G	Val Val Val Val	Ala Ala Ala Ala	Asp Asp Glu Glu	Gly Gly Gly Gly	U C A G

Aminokyseliny

Obrázek 4: Genetický kód. Umístění bází v první, druhé a třetí pozici tripletu. Zkratky (Phe, Leu atd.) označují aminokyselinu kódovanou daným tripletem (viz: <www.genome.gov>).

Pojmenování vztahu bází řazených v mRNA a aminokyselin jakožto genetického kódu se vztahuje k jednomu z klasických pojmů lingvistiky. Pojem kód vyjadřuje vztah dvou veličin daný určitým územ (Monod 1970, s. 159–160). Od pojmenování genetický kód se konzistentně odvíjí i další názvosloví – kromě samotného kódování aminokyseliny tripletem (kodonem) je celý proces tvorby aminokyselinového řetězce podle nukleotidového vzoru při proteosyntéze pojme-

nován jako translace. Ta se staví do opozice vůči jednoduchému přepisu vzájemně chemicky afinitních molekul při transkripci.

Lingvistická paralela

V instrumentáriu jsme popsali strukturu DNA a proces proteosyntézy. V souvislosti s tím jsme představili také tradiční molekulárně-biologickou terminologii. Tato terminologie často využívá lingvistických termínů (transkripce, translace, kód, text) nebo termínů založených na jazykové metafoře (zápis, čtení, zpráva, informace (neterminologicky)). Jakobson (1971) potvrzuje, že využití jazykové metafory v molekulární biologii je korektní a že genetický kód má vlastnosti přirozeného jazyka. Analogizuje báze s písmeny (respektive fonémy), triplety se slovy a geny s větami. Nachází mnohé další společné vlastnosti genetického kódu a přirozeného jazyka. Ji (1999, s. 412) postupuje dále a analogizuje širokou škálu vlastností genetického kódu a přirozeného jazyka: písmena s nukleotidy a aminokyselinami, slova s geny, řetězce slov se souborem společně exprimovaných genů. Dále k sobě vztahuje gramatiku a fyzikální a chemické zákony, fonetiku a řízení energetického toku, sémantiku a genově řízené procesy v buňce. V případě obou kódů explicitně hovoří o dvojí artikulaci. Ji v nalézání protějšků procesů v buňce a konceptů popisujících přirozený jazyk představuje extrémní případ. Jeho přístup sugeruje, že libovolnému lingvistickému konceptu lze nalézt odpovídající proces či strukturu v buňce.

Trifonov (1988) popisuje soustavy kódů zajišťujících interakci DNA, RNA a proteinů, u jiných autorů můžeme nalézt další obdobné metafory (viz např. Barbieri 2002; Collado-Vides 1992; 1993; Markoš 1997; Markoš 2002).

Diskurz těchto znakových popisů procesů v buňce je rozvíjen biosémiotikou, mladou vědní disciplínou. Problém současné biosémiotiky spočívá v tom, že výše zmíněné a mnohé další znakové přístupy k buňce jsou vzájemně nekonzistentní. Každý autor rozvíjí specifický přístup a neexistuje jednotná metoda, která by platnost těchto přístupů ověřovala. Na příkladu Jiho lze vidět, že analogizovat přirozený jazyk

a genetický kód lze libovolně, přičemž posouzení korektnosti takových analogií není snadné.

Motivace experimentu

Metafory a analogie nám mohou poskytnout nadhled nad určitou problematikou. To je ale v kontextu jazykových metafor a analogií DNA problematické. Nevíme, které z těchto metafor jsou relevantní a užitečné a které nikoliv. Popis procesů v buňce využívá jazykovou metaforu a analogii. Uvažování genetiků, bioinformatiků, makromolekulárních biologů, biochemiků a dalších tak může být ovlivněno zavádějící metaforou. Z epistemologického hlediska by korekce těchto metafor měla pro jejich uživatele velký význam. Cílem tohoto článku je představit experimentální metodu, která by mohla ověření některých ze zmíněných metafor umožnit a zároveň poskytnout vhled do struktury genetického kódu.

Metodika kvantitativní analýzy DNA

Pro analýzu nukleotidových sekvencí reprezentovaných zápisem v textu jsou užívány různé lingvistické metody a kvantitativní analytické přístupy. Představíme některé z nich a ukážeme, jak se vztahují k naší metodě analýzy struktury genetického kódu. Pokusíme se o využití těchto metod pro podložení nového designu genetického kódu.

Mantegna et al. (1995; viz též Havlin et al. 1995) analyzovali projevy Zipfova zákona na kódující a nekódující DNA. Kódující DNA dle Mantegni et al. Zipfův zákon vykazuje. Nekódující DNA projevy Zipfova zákona vykazuje také, ale pouze do určité míry. Mantegnova analýza byla motivována poznatkem, že pouze malá část (pro homo sapiens uvažováno 5,33 %; Mantegna et al. 1995, s. 2940) genomu je kódující a tedy nese informaci k výstavbě proteinu (viz naše instrumentarium výše). Zbylá část genomu nemá takovou jasně zadanou funkci a od šedesátých let se pro ni zažil termín junk DNA. Nekódující DNA měla

být v genomu historicky neseným reliktem bez využití v organismu (viz např. Watson – Berry 2003, s. 253; Palazzo 2014; Mantegna et al. (1995) o junk DNA píše jako o silent DNA).

Mantegna et al. (1995, s. 2949) dále tvrdí, že se nekódující DNA podobá v některých vlastnostech přirozenému jazyku (viz též Niyogi – Berwick 1995). Mantegna et al. (1995) mluví o tom, že nekódující DNA nese určitý jazyk, z hlediska jeho redundance oproti kódující DNA dokonce bližší přirozeným jazykům (tím Mantegna et al. rozšířili analogie DNA a přirozeného jazyka, o nichž hovoříme níže, a opět uplatnili jazykovou metaforu DNA). Tato zjištění Mantegnu et al. vedou k hypotéze, že nekódující DNA má také funkci, kterou prozatím neregistrujeme a nepopisujeme, a tedy že nekódující DNA je nějakým způsobem použita pro uchování informací „biologických struktur“ (Mantegna et al. 1995, s. 2949). Pozdější rozvoj molekulární biologie dal Mantegnově domněnce za pravdu (viz Alberts et al. 2008, s. 31-42; The ENCODE Project Consortium 2012, s. 57).

Potvrzení výskytu Zipfova zákona u kódující DNA odpovídalo tomu, že kódující řetězec nese informaci k výstavbě funkčního tvaru proteinu, tj. určité struktury s určitou funkcí v organismu. Analogicky k tomu se v textech v přirozeném jazyce projevuje Zipfův zákon z důvodů naplňování určité funkce textu (to se můžeme pokusit vysvětlit např. v souvislosti s informační strukturou textu zajišťující přenos signálu prostředím a výrazovou a obsahovou strukturou a soudržností textu; viz Zipf 1949, s. 19–47).

K Mantegnově et al. analýze je ale nutné poznamenat následující: V analýze Mantegna et al. využívají dlouhé řetězce nekódující DNA (delší než 50 tisíc bází). Nekódující DNA sice informaci nese, ale nese také množství reliktních řetězců bez využití (nekódují protein ani se nepodílejí na regulaci proteosyntézy). Projekt ENCODE (2012) odhaduje, že až 80 % nekódující DNA má funkční využití. Zbýlých 20 % muselo Mantegnovu et al. analýzu ovlivnit, a to proto, že v jeho analyzované nekódující DNA musely být obsaženy složky regulace proteosyntézy a také reliktní DNA (pro niž můžeme stále používat termín junk DNA a která obsahuje mnoho repetitivních a z informačního hle-

diska redundantních sekvencí). Tato kontaminace by pak posilovala hodnocení nekódující DNA jako podobné přirozenému jazyku z hlediska redundance.

Mantegna et al. byli při zacházení s nekódující DNA postaveni před následující problém: koncept genetického kódu přisuzuje kódující DNA strukturální roli tripletů (viz vztah tripletů a aminokyselin popsaný výše). Nekódující DNA však takové striktní ohraničení a priori přisuzovat nelze, funkční roli zde mohou zastávat sekvence o různé délce. Proto se pro kvantitativní analýzu kódující i nekódující DNA rozhodl využít techniku tzv. n-gramů (tuto techniku využívají např. také Bolshoy et al. 2010, s. 26). Představíme ji na ilustračním příkladu.

Mějme následující řetězec ABCDEFGH. Tento řetězec segmentujeme na 3-gramy, jimiž jsou: ABC, BCD, CDE, DEF, EFG, FGH; 4-gramy mají podobu: ABCD, BCDE, CDEF, DEFG, EFGH. N-gramová analýza tedy segmentuje řetězec tak, že postupuje lineárně jednotku po jednotce a delimituje vždy v řetězci následující n-tici (n-gram). V analýze přirozených textů n-gramová segmentace postupuje bez registrace hranic slov či vět. Sekvence „Zipfův zákon“ je rozdělena na tyto 5-gramy hlásek: zipfů, ipfův, pfůvz, fůvzá, ůvzák atd. Při analýze kódující i nekódující DNA jsou podobně zanedbány jakékoliv dříve stanovené hranice. N-gramový přístup tak dovoluje analyzovat řetězec nezávisle na jeho vnitřní struktuře tím, že registruje jednotlivé sousedící prvky. Jednotlivé n-gramy představují v analýze analogie slov, která nejsou vydělena mezerou, ale hranicí délky n-gramu (zipfů, ipfův, pfůvz atd. představují slova vstupující do analýzy).

Tento přístup zvolili k oběma typům DNA Mantegna et al. My tímto způsobem postupujeme také, a to z toho důvodu, abychom se vyhnuli apriornímu určení hranic kódovaných složek řetězce (viz dále). Jednotky delimitované n-gramovou technikou budeme stejně jako Mantegna et al. analyzovat z hlediska projevu Zipfova zákona. Výsledky analýzy nás mají vést k potvrzení aktuálního designu genetického kódu, nebo případně k jeho odmítnutí a následné reformulaci.

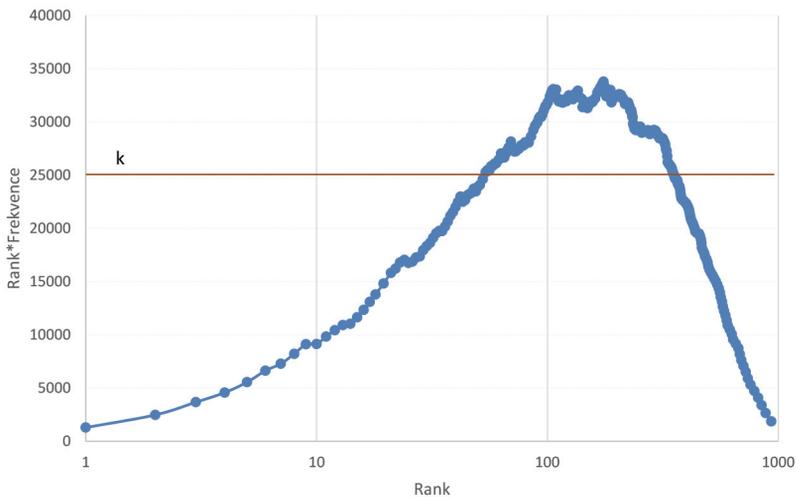
Je však nutné poznamenat, že využití Zipfova zákona v naší analýze může být sporné. Kvantitativní charakteristiky textu, jako jsou projevy

Zipfova zákona nebo n-gramové analýzy, mohou být jako důkaz o určité vlastnosti či povaze tohoto textu (je kódující / nese informaci / nese instrukci; znaková funkce v genetickém kódu) chápány jako nepřímé. Např. Konopka (1995) referuje o závěrech Mantegna týmu a doplňuje poznámky, které mohou být analogicky vztaženy i k našim závěrům. V prvním případě vyvstává problém s tím, že projevy Zipfova zákona jsou identifikovány na mnoha jevech, jako je trh, velikost měst nebo biologických populací atd. Zipfův zákon se pak jeví jako epifenomen jakéhokoliv systémového chování fyzikálních, sociálních, biologických apod. soustav. Zipfův zákon se tedy nemusí vztahovat ke kódujícím funkcím řetězce, ale naopak k jiným jevům jeho konstrukce. V případě řetězce báží by to mohla být jejich kombinatorika daná termmostabilitou jednotlivých báží. V případě přirozených textů by se mohlo jednat např. o fonetické důvody kombinovatelnosti hlásek, které představují typ systematizace.

Zipfův zákon byl dokonce napadán v obecném rozměru, a to na základě jeho projevů na náhodných a nenáhodných textech (Li 1992); diskuse v této oblasti zahrnuje problém generování náhodného vzorku a to, že mechanismus tvorby náhodného vzorku může produkovat projevy, které se z důvodů ne zcela náhodného generování textu přiblíží Zipfovou zákonu – to by opět potvrdovalo domněnku, že Zipfův zákon je projevem libovolného systémového, auto/regulovaného chování). Stále je ale vnímán jako projev znakovosti, kódovosti či jazykovosti. Na základě něj se hodnotí např. dorozumívání zvířat nebo struktura textu pacientů s postižením způsobujícím jazykový deficit (Ferrer-i-Cancho 2006; Ferrer-i-Cancho – McCowan 2009; Ferrer-i-Cancho – Elvevag 2010). I přes veškeré zmíněné výhrady použijeme v naší analýze Zipfův zákon, navážeme tak na lingvisticko-kvantitativní diskurz ověřování kódové povahy DNA, jako je tomu u Mantegni et al. a dalších. K tématu Zipfova zákona a DNA viz Tsonis et al. 1997.

Zipfův zákon na přirozených textech

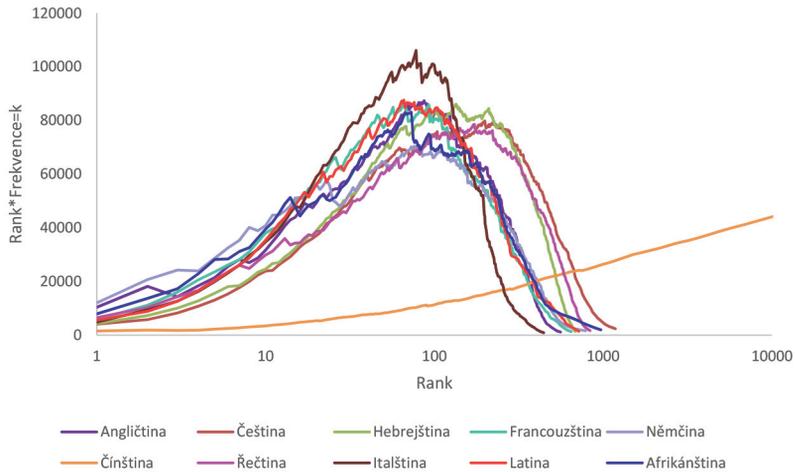
Zipfův zákon formuluje následující vztah: vezmeme-li určitý text a seřadíme-li počty výskytů (neboli frekvence) jeho entit (např. slov) od nejvyšší po nejnižší, pak pokud vynásobíme frekvenci každé této entity jejím pořadím (tzv. rankem), bude se výsledek p vždy blížit určité hodnotě reprezentované tzv. konstantou k (Zipf 1949, s. 22–25). Jak si ale můžeme všimnout na obrázku 5, výsledky násobků ranků a frekvencí jsou u přirozených jazyků značně proměnlivé. Pomocí vodorovné úsečky proto v obrázku ilustrativně zobrazujeme konstantu k , jak ji vyjadřuje Zipfův zákon – reálné hodnoty násobků ranků a jejich frekvencí jsou od ní různě vzdáleny.



Obrázek 5: Vztah ranku a násobku ranku a frekvence písmen českého textu a ilustrace konstanty k Zipfova zákona. Text: M. Viewegh – Účastníci zájezdu.

Experimentální metoda založená na Zipfově zákonu spočívá ve sledování průběhů grafů hodnot násobků ranků a frekvencí entit textu (tj. spočívá ve vizuální komparaci průběhů grafů a registrování jeho vlastností, např. konkávnost, konvexnost, strmost, charakter maxim, defi-

niční obory, monotónnost atd.; prozatím jsme neaplikovali žádnou formální metodu vyjádření podobnosti grafů, použitá kritéria jsou však pro naše účely dostačující). Tyto entity vybíráme z jedné konkrétní jazykové roviny textu – písmena, slova, věty apod. Klíčovým aspektem této metody a naší experimentálně ověřenou tezí je, že se u různých přirozených jazyků jednotky konkrétních jazykových rovin (písmena, slova, věty apod.) projevují podobně. Máme-li pak text v neznámém jazyce či neznámém zápisu, můžeme díky této metodě identifikovat jazykovou povahu jeho jednotek. Připomínáme jen, že je při této analýze využita n-gramová technika. Příklad uvádíme na obrázku 6. Na něm můžeme sledovat projevy Zipfova zákona u deseti různých jazyků. Studovanými jednotkami jsou zde dvojice písmen textu (2-gramy; nejsou registrovány žádné spřežky, pracuje se s nimi jako s kombinací jednotlivých písmen, např. spřežka ch. Tečky, mezery, pomlčky atd. nejsou registrovány, registrována jsou pouze písmena). Všechny texty byly před analýzou redukovány na stejný počet znaků (270 000). Čeština, angličtina, němčina, afrikánština, latina, italština, francouzština, řečtina a hebrejštiny mají obdobný průběh grafu, čínština se od ostatních průběhů výrazně liší. Z tohoto pozorování můžeme usuzovat, že znaky čínštiny mají oproti ostatním zkoumaným jazykům zcela jinou jazykovou roli.



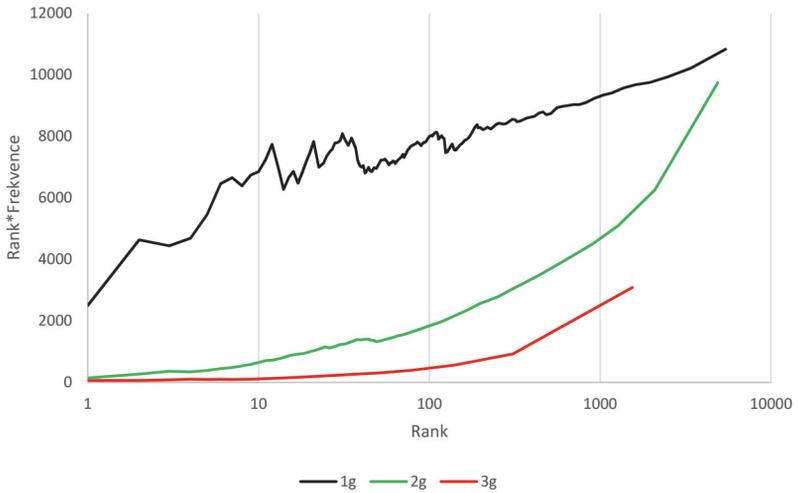
Obrázek 6: Projev Zipfova zákona na 2-gramech písmen textů vybraných jazyků. Texty: afrikánština <18203-8>; řečtina <28658-0>; hebrejština <8cewa10>; latina <27219-0>; němčina <30695-0>; italština <28910-8>; francouzština <15371-8>; čínština <25350-0>; angličtina J. R. R Tolkien – The Lord of the Rings; čeština M. Viewegh – Účastníci zájezdu. Kód ve špičatých závorkách je identifikátor textu volně stažitelného v projektu Guttenberg (<www.gutenberg.org>).

V předchozím odstavci jsme užili naši experimentální metodu na vzorku textů různých jazyků segmentovaných v tomto případě na 2-gramy písmen. Náš experimentální postup však provádí stejnou analýzu za užití 1-gramů, 2-gramů, 3-gramů atd. (experimentálně jsme ověřili signifikantnost nejvýše 20-gramů písmen). Tento způsob analýzy nám umožňuje sledovat, jak postupně se zvětšující n-tice odrážejí strukturu textu, a to v určité kontinuitě proměn podoby grafu. Postupné zvětšování n-tic nám dává možnost sledovat strukturu textu na stále vyšších jazykových rovinách. Pozorujeme tak chování písmen (jakožto stanovené základní úrovně popisu), kombinací písmen, slov (daných jejich průměrnou délkou v určitém jazyce, pro češtinu je to 5-gram písmen; viz dále), dále vět atd. Všechny tyto jednotky jsou reprezentovány jako n-gramy písmen. Tento experimentální postup je plausibilní, v mnoha analýzách jsme ověřili, že průběhy grafů n-gramů

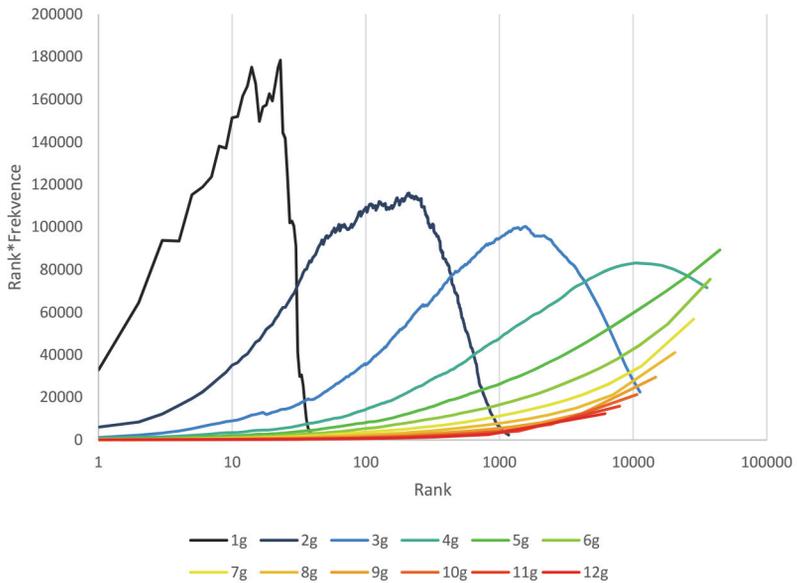
odpovídajících průměrné délce dané jednotky v určitém jazyce (např. slov) jsou signifikantně podobné průběhům grafů těchto jazykových jednotek textu. Tuto proceduru nazýváme mapování.

Výsledkem procedury mapování určitého textu je x průběhů (zobrazujeme je do jednoho grafu), které můžeme použít pro porovnání s výsledkem procedury mapování jiného textu. Porovnáním se mívá zjištění podobnosti jednotlivých navazujících n -gramových průběhů u obou textů. Například pokud známe jazykové jednotky určitého textu, který dále zmapujeme, můžeme použít výsledný graf jako referenci k porovnání s výsledným grafem jiného textu, u kterého neznáme povahu jeho jazykových jednotek. Tímto způsobem se můžeme pokusit o jejich určení. Konkrétní příklady srovnání českého a čínského textu uvádíme níže.

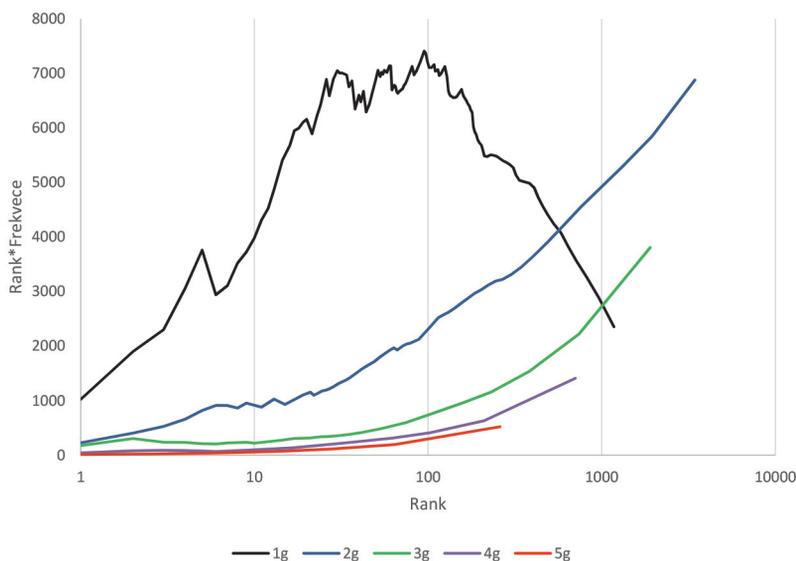
První graf (obrázek 7) ukazuje mapování slovních forem českého textu. Druhý graf (obrázek 8) ukazuje mapování písmen rovněž českého textu. Na základě korelace jejich průběhů zjišťujeme, že průběh 1-gramů slov se podobá průběhu 4-gramů až 5-gramů písmen. Průběhy následujících n -gramů (u slov 2-gramy a vyšší, u písmen 5-gramy a vyšší) si svými průběhy také odpovídají. Korektnost korelace 1-gramů slov s 4-gramy a 5-gramy písmen je zajištěna také podobností vývoje grafů v obou mapováních, nikoliv pouze podobností dvou konkrétních průběhů grafů. K nalezené hranici 4-gramů až 5-gramů písmen můžeme poznamenat, že průměrná délka českého slova, jak jsme experimentálně zjistili (na vzorku 243 českých literárních textů velikosti od 140 slovních forem do 149 070 slovních forem), má hodnotu 4,768 písmen. Srovnání obou uvedených mapování této hodnotě odpovídá, n -gramová analýza identifikuje hranici počtu písmen odpovídající slovům.



Obrázek 7: Mapování slovních forem českého textu. Text: M. Viewegh – Účastníci zájezdu.



Obrázek 8: Mapování písmen českého textu. Text: M. Viewegh – Účastníci zájezdu.



Obrázek 9: Mapování znaků čínského textu. Text: čínština <25350-0>. Kód v ostrých závorkách je identifikátor textu volně stažitelného v textu Guttenberg (<www.gutenberg.org>).

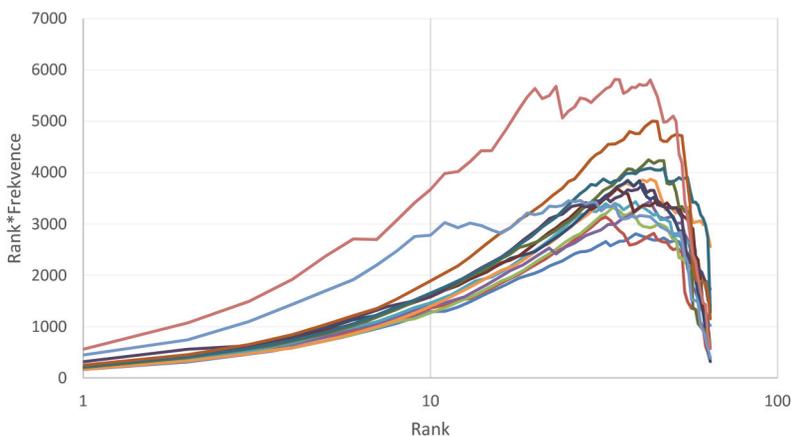
V obrázku 9 je zobrazeno mapování znaků čínského textu. To srovnáme s předchozím mapováním písmen českého textu (obrázek 8). Sledujeme-li průběhy grafů mapování čínského textu, identifikujeme podobnost průběhu 1-gramů znaků čínského textu s 2-gramy až 3-gramy písmen českého textu. Podobně 2-gramy (a vyšší) čínského textu pak odpovídají průběhu grafů 5-gramů (a vyšších) písmen, respektive 2-gramů slov českého textu. Zjišťujeme tak, že znaky čínského textu nemají povahu písmen, ale spíš jejich dvoj až trojkombinací, což odpovídá charakteru čínského znakového písma. Na základě výše stanovených kritérií porovnání průběhů grafů je korektnost tohoto závěru opět potvrzena.

Ověření metafory DNA

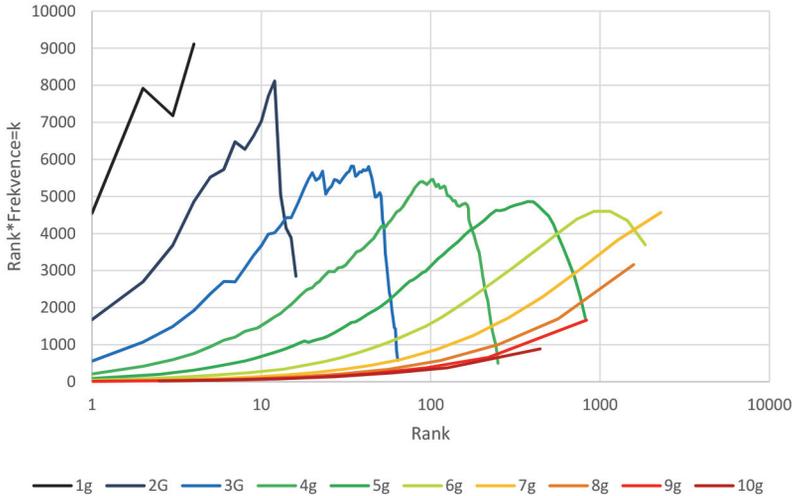
Výše uvedenou metodou mapování budeme analyzovat sekvence mRNA. Představili jsme standardní pojetí genetického kódu a různé analogie DNA a jazyka, včetně společných metod jejich zkoumání. Nejužívanější z analogií DNA a jazyka je pojetí bází DNA jakožto písmen – následně tripletů jakožto slov a genů jakožto vět (Jakobsonova metafora). Aplikací naší metody mapování vzorků sekvencí mRNA chceme ve srovnání s mapováním textů přirozeného jazyka prověřit, zda báze DNA a konsekventně mRNA hrají ve struktuře genetického kódu obdobnou roli jako písmena ve struktuře přirozených jazyků.

Aplikace metody

Výše představenou metodou bylo zmapováno přibližně 1000 náhodně vybraných mRNA sekvencí homo sapiens (EMBL-EBI, 2014; použité mRNA sekvence mají různou délku, experimentálně jsme ověřili, že délka textu nemá vliv na mapování, pouze na obor hodnot, který není kritériem porovnání grafů). Zvolenou jednotkou je v této analýze báze. Aplikací naší metody (testováno na 1-gramech až 20-gramech) jsme zjistili signifikantní podobnost grafů různých sekvencí (viz např. obrázek 10 s grafy 3-gramů). Pro názornost grafu mapování však uvádíme analýzu pouze jedné náhodně vybrané sekvence (obrázek 11). Pro přehlednost ještě zopakujeme, že obrázek 10 ukazuje 3-gramovou analýzu Zipfova zákona více vzorků mRNA a obrázek 11 1-gramovou až 10-gramovou analýzu Zipfova zákona jedné konkrétní sekvence.



Obrázek 10: Projev Žipfova zákona 3-gramů bází v 20 náhodně vybraných sekvencích mRNA z analyzovaného vzorku. Text: ENA <AAA59187> 1 Homo sapiens (human) ras GTPase-activating-like protein, ENA <AAA59483> 1 Homo sapiens (human) epiligrin alpha 3 subunit, ENA <AAA59486> 1 Homo sapiens (human) laminin B1, ENA <AAA60554> 1 Homo sapiens (human) sodium channel alpha subunit, ENA <AAA59504> 1 Homo sapiens (human) lactase phlorizinhydrolase, ENA <AAA18895> 1 Homo sapiens (human) voltage-gated sodium channel, ENA <AAA51901> 1 Homo sapiens (human) calcium channel L-type alpha 1 subunit, ENA <AAA35629> 1 Homo sapiens (human) calcium channel alpha-1 subunit, ENA <AAA51898> 1 Homo sapiens (human) N-type calcium channel alpha-1 subunit, ENA <AAA15448> 1 Homo sapiens (human) DNA polymerase epsilon catalytic subunit, ENA <AAA60225> 1 Homo sapiens (human) protein tyrosine phosphatase zeta-polypeptide, ENA <AAA18639> 1 Homo sapiens (human) p300 protein, ENA <AAA59866> 1 Homo sapiens (human) mannose 6-phosphate receptor, ENA <AAA59924> 1 Homo sapiens (human) GAP-related protein, ENA <AAA58965> 1 Homo sapiens (human) collagen type VII, ENA <AA52700> 1 Homo sapiens (human) heparan sulfate proteoglykan. Kód v ostrých závorkách je identifikátor textu volně stažitelného v genové bance EMBL-EBI (<www.ebi.ac.uk>).



Obrázek 11: mRNA. Text: ENA <AAA52700> 1 Homo sapiens (human) heparan sulfate proteoglykan. Kód v ostrých závorkách je identifikátor textu volně stažitelného v genové bance EMBL-EBI (<www.ebi.ac.uk>).

Věnujme se nyní obrázku 11, který porovnáme s obrázkem 8 zobrazujícím mapování písmen českého textu. Můžeme si všimnout, že průběhy grafů mapování jsou si podobné od 3-gramů bází u DNA a 1-gramů písmen českého textu. Průběh 2-gramů bází je vůči průběhu 1-gramů písmen neúplný a až 3-gram bází realizuje křivku, kterou nacházíme u 1-gramů písmen (viz obrázek 12 zobrazující mapování distinktivních rysů hlásek, o kterých hovoříme níže – i zde je rozhodujícím kritériem úplnost průběhu grafu a podobnost průběhu grafu jako taková). Z tohoto důvodu klademe hranici písmene k 3-gramům bází. Následující průběhy obou mapování mají podobný vývoj. Průběhy 1-gramů a 2-gramů bází DNA nejsou u písmen realizovány, 3-gramy bází odpovídají 1-gramům písmen, 6-gramy až 7-gramy bází odpovídají 4-gramům až 5-gramům písmen. Další průběhy mají totožnou povahu.

Z průběhu obou mapování můžeme implikovat následující. Báze jsou jednotkou konstitučně nižší než písmena. U 2-gramů bází můžeme sledovat podobný, avšak neúplný průběh, jako mají 1-gramy hlá-

sek (viz obrázek 12 zobrazující mapování distinktivních rysů hlásek, kde totéž platí pro 1-gramy až 5-gramy distinktivních rysů). To je vysvětlitelné degenerovaností genetického kódu, kde je pro kódování aminokyseliny často třetí báze redundantní (obdobně mnoho hlásek odlišují pouze jeden či dva distinktivní rysy). Trojici bází (tripletům) přisuzujeme na základě podobnosti průběhů grafů roli písmen, sedmice bází (tedy více než 2 triplety) již tvoří obdobný typ konstituentu jako čtyř až pětice písmen – tj. tvoří obdobu slov.

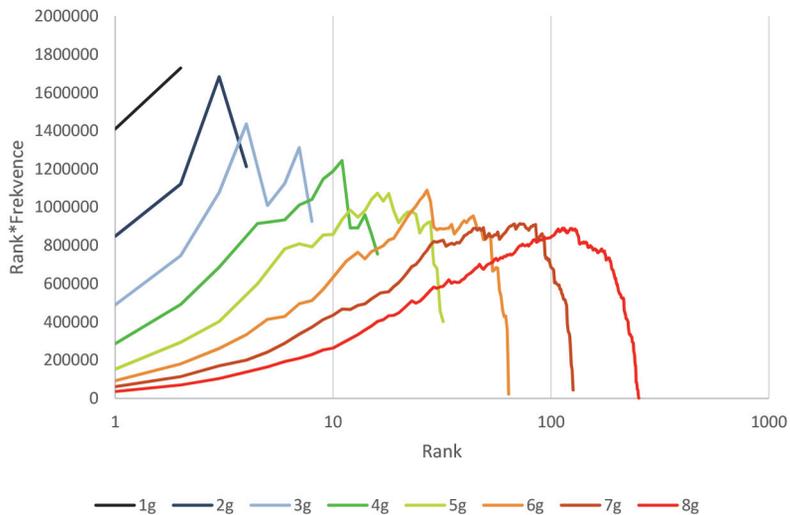
Vezměme první implikaci, která říká, že báze jsou jednotkou konstitučně nižší než písmena. Z této implikace vyplývá otázka, jakou roli hrají báze v genetickém kódu, analogizujeme-li jej s přirozeným jazykem a předpokládáme-li obdobný design obou kódů.

Písmena zcela intuitivně vnímáme jako nedělitelná. Tvořena jsou ovšem na základě vzájemných vztahů, které diferencují jedno písmeno od druhého. Každé písmeno je pak tvořeno souborem vlastností, který jej charakterizuje a zároveň odlišuje od ostatních. Lingvistika tyto vlastnosti nazývá distinktivními rysy. Písmena textu tak můžeme chápat jako soubory distinktivních rysů. S touto rovinou jsme při analýze přirozeného jazyka prozatím nepracovali.

Pro zavedení roviny distinktivních rysů do analýzy postupujeme následujícím způsobem: každé z písmen charakterizujeme jeho vlastnostmi (distinktivními rysy), které jej odlišují od ostatních. Neužili jsme tradičně lingvistikou popisované distinktivní rysy ve smyslu akustických vlastností hlásek, kterých je ve fonologických popisech více než deset. Použili jsme nejmenší možný počet k písmenům arbitrárně přiřazených distinktivních rysů schopných odlišovat písmena češtiny. Pro českou abecedu je takových opozic nutných pouze šest. V textu je každé písmeno reprezentováno unikátním řetězcem šesti pozic obsazených jedničkou nebo nulou (tedy přítomných nebo nepřítomných vlastností arbitrárních distinktivních rysů). Takto nově reprezentovaný text zmapujeme a porovnáme s výsledky mapování mRNA.

Výsledek mapování distinktivních rysů českého textu (viz obrázek 12) nám odhaluje typy průběhů, které se projeví u 1-gramů a 2-gramů bází DNA (obrázek 11). U mapování písmen podobné průběhy

nenacházíme. Průběhům 1-gramů až 5-gramů distinktivních rysů písmen českého textu ovšem odpovídají průběhy grafů 1-gramů a 2-gramů bází. U distinktivních rysů je z důvodu jejich počtu mezi průběhy pozvolnější přechod. Průběhu grafu 6-gramů distinktivních rysů odpovídá průběh 3-gramů bází. Z toho můžeme usuzovat, že báze mají v designu genetického kódu obdobnou roli jako distinktivní rysy písmen v designu přirozeného jazyka.



Obrázek 12: Mapování distinktivních rysů českého textu.

Interpretace výsledků experimentu

Na základě aplikace Zipfova zákona na n -gramy bází DNA, písmena textu přirozeného jazyka a distinktivní rysy písmen textu přirozeného jazyka můžeme usuzovat, že nukleotidové báze DNA plní v designu genetického kódu roli analogickou distinktivním rysům písmen textu přirozeného jazyka. Tripletty bází jsou dále analogické písmenům. Kombinace tripletů jsou následně analogií slov.

Naším cílem bylo ověření analogie DNA a přirozeného jazyka. Jakobson formuloval analogii bází DNA a písmen, tripletů a slov, genů a vět. Naše analýzy tento design genetického kódu zpochybňují a ukazují analogii bází DNA s distinktivními rysy, tripletů s písmeny a kombinací tripletů se slovy. Všeobecně užívaná analogie bází jakožto písmen genetického textu se tak jeví jako chybná. Reprezentace genetického zápisu sledem písmen odpovídajících bázím (A, C, G, T) zřejmě zapříčinila pevné ukotvení této analogie ve vědecké praxi a v laickém pojmání genetického kódu. Naše výsledky však jasně ukazují, že tato analogie je vzhledem k designu přirozeného jazyka nesprávná, a že „báze nejsou písmena“.

Konstruování gramatik DNA

V této kapitole ukážeme přístupy ke gramatickému popisu DNA a proteosyntézy, které doposud nebyly příliš reflektovány a které lze z hlediska typologie gramatik charakterizovat jako gramatiky postupující od obsahu k výrazu. Východiskem a oporou naší úvahy je práce Jana Kořenského *Konstrukce gramatiky ze sémantické báze* (1984). Návrh gramatiky představený v tomto Kořenského textu se nabízí k použití při popisu sémiotických procesů v buňce či živém tvaru. Při záměně přirozeného jazyka za DNA a proteosyntézu naznačuje Konstrukce gramatiky ze sémantické báze postupy, jakými by mohlo být zacházeno s jejich znakovým obsahem. Činí tak s vymezením vztahu obsahu k pragmatické dimenzi a ke komplexní výrazové struktuře znaku. Vysvětlíme, co znamená formulace konstruování gramatiky od výrazu a konstruování gramatiky od obsahu. Představíme, jaké aspekty s sebou nese konstruování gramatik od výrazu a naopak od obsahu a v čem jsou jeden i druhý přístup problematické. Zaměříme se také na to, jak lze obě hlediska kombinovat. Reagovat při tom budeme na některé gramatické modely DNA a proteosyntézy, které explicitně popisují znakové procesy v živém tvaru.

Co je to gramatika

Gramatikami v nejširším možném slova smyslu můžeme rozumět systematizované popisy znakových projevů. Znakovým projevem je například slovo i komplexnější jazykový útvar, věta/výpověď nebo text. Dále mimika a gestika reflektovaná v komunikaci. Znakovým projevem je určitý typ chování, oděv, účes a každý předmět, jenž užíváme. Znakovým projevem je umělecké dílo, rozvržení určitých prvků v architektuře, náš otisk prstu, který jsme někde nevědomě zanechali. Třída jevů, které můžeme charakterizovat jako znak, je značně rozmanitá. Co vše a jak se projevuje jako znak, naznačují obsahy přehledových sémio-

tických publikací či sémiotické studie orientované na specifická témata (viz Nöth 1990). Můžeme každopádně jasně stanovit, co za znakový projev považovat lze, a co nikoliv.

Z hlediska přístupu k popisu znakových projevů můžeme odlišovat a typologizovat gramatiky na dva typy. Gramatikou může být myšlen systematizovaný popis znakových projevů, u kterých jsou předpokládáni dva a více účastníci. Stejně tak může gramatika popisovat reakce jedince vnímajícího vnější svět (tzv. indexické znaky), kdy se sémiózy bezprostředně účastní jen jeden subjekt. Za různé gramatiky jsou dále považovány modely předpokladů produkce a modely předpokladů recepce jisté třídy textů. Gramatiky mohou postihovat komunikační interakce nejrůznějších podavatelů a příjemců sdělení v typizovaných situacích, mohou se vztahovat k jediné komunikační události a formulovat pravidla, která její subjekty postupně utvářejí (Viz Kořenský 1984, 1987, 1988, 1992, 1994, 1996a, 1996b, 1997, 1999, 2000, 2003, 2006). Nejčastěji jsou ale gramatikami míněny vědecké popisy národních jazyků.

Gramatický popis DNA je od svého původu spjat s jazykovou metaforou a s analogií k dobovým mluvnicím a fonologickým popisům národních jazyků. Gramatické instrumentarium bylo na DNA a proteosyntézu přenášeno z rovinových lingvistických popisů národních jazyků vytvořených na začátku druhé poloviny dvacátého století (až později byly přenášeny i generativní modely viz Collado-Vides 1992, 1993; Searls 2002). Sémiotické reflexe molekulárně-genetického popisu přímo analogizovaly jazykové plány a strukturu DNA a proteinů, jednotky určitých plánů. To jsme v předchozí kapitole knihy ukázali, včetně revize jednotkového pojetí bází. Explicitní či implicitní gramatický popis DNA a proteosyntézy se od svého počátku nerozvíjel nijak systematicky. Nedochovalo k vymezení popisu výrazové struktury na jedné straně a obsahové na druhé. Výraz ani obsah DNA a proteosyntézy nebyl definován. Molekulárně-biologický výzkum rozvíjel poznání DNA či proteosyntézy a sémiotické aspekty obou jevů se spontánně podřizovaly analogii s národním jazykem. Pouka-

zuje na to Raibleho šetření (2001), které jsme zmínili v předchozí kapitole.

Formulování gramatik DNA a proteosyntézy je bez ohledu na historické aspekty oprávněné. Znaková povaha těchto jevů zajišťuje, že můžeme hovořit o jejich gramatickém popisu. Texty národních jazyků a DNA mají dvě hlavní společné vlastnosti, linearitu a arbitrárnost. A právě arbitrárnost je jakýmsi identifikačním minimem znakové povahy. V případě DNA a proteosyntézy se jí míní princip přiřazení tripletu k aminokyselině, řetězu bází DNA k sledu aminokyselin v proteinu.

„(...) fakt, že už je kód rozluštěný a ví se, že je univerzální, nám alespoň dovolu je formulovat problém přesně, jen s nepatrným zjednodušením: (a) Buď strukturu kódu objasňují chemické – nebo přesněji – stereochemické důvody, tj. byl-li ‚zvolen‘ určitý kodon pro zastoupení aminokyseliny, pak proto, že mezi nimi existovala jistá stereochemická afinita; anebo (b) je struktura kódu chemicky libovolná: kód, jak ho dnes známe, je výsledkem řady nahodilých voleb, jež ho postupně obhacovaly. První z těchto hypotéz se zdá daleko svůdnější. Za prvé proto, že by vysvětlovala univerzalitu kódu. Dále, protože nám dovolu je představit si primitivní překladový mechanismus, v němž by sekvencí ražení aminokyselin do polypeptidu bylo způsobené přímou interakcí mezi aminokyselinami a samotnou replikativní strukturou. Konečně a především proto, že by tato hypotéza, pokud by byla pravdivá, byla v zásadě ověřitelná. V tomto směru bylo už vyvinuto hodně snahy, zatím bezvýsledně (...) biosféra se samou univerzalitou svých struktur, kódem počínaje, zdá být výsledkem události jedinečné.“ (citace, s vynecháním odkazů, Monod 2008, s. 136-137, překlad A. Markoš)

Zastavme se nyní u výrazu a obsahu, které mají arbitrární vztah. V přirozených jazycích rozeznáváme zvuky o jisté délce, výšce či charakteru šumu, s daným vztahem výšek tónů, s průběhem výšky základního tónu atd. Tyto zvuky (fyzikální podmínky užití v systému distinkcí) jsou konstrukčním prostředkem pro hlásky určitého jazyka.

Část fyzikálního světa je tak vzata a implementována v sémióze jako výraz, který je spojován s obsahy. Ani obsahová rovina nepracuje s celým fyzikálním světem – do funkce s výrazy vstupují jen jisté kvality fyzikálního světa, které můžeme vnímat. To, že je vztah výrazu a obsahu arbitrární je dáno ekonomii kódu. Pokud by význam vyžadoval konkrétní výraz, a naopak, nebyla by sémióza ve fyzikálním světě stabilní. To jsme ukázali v první kapitole na příkladu z Barnibalbi. O výrazu a obsahu hovoříme jako o dvou sadách, které jsou k sobě vztaženy ve znakové funkci. Ve výše uvedené citaci z Monoda je formulována arbitrárnost vztahu DNA a proteinů. Procesy, které jsou založeny na dvou arbitrárně spojených sadách, jsou popisovány gramatikami. Tato argumentace je nutná, aby bylo objasněno užívání spojení gramatika DNA.

Jakobsonovy analogie jazyka a DNA

Vraťme se k Jakobsonově analogii DNA a přirozeného jazyka, kterou jsme popsali v předchozí kapitole. Jakobsonův rovinový popis DNA a proteinů analogizuje báze a písmena (respektive fonologický systém), triplety a slova (lexikální systém), geny a věty (syntaktický systém). Jakobson registruje požadavek arbitrárního vztahu výrazu a obsahu a předestírá velmi jednoduchý koncept gramatiky DNA. Z jeho výkladu lze rekonstruovat, že výrazem je DNA a obsahem proteiny.

Systém bází je analogický s fonologickým systémem, to je výchozí moment Jakobsonova gramatického modelu. Kombinatorika v DNA protilehlých purinů a pyrimidinů, při níž adenin vždy stojí proti thyminu a nikoliv proti adeninu a guaninu – podobně to platí pro ostatní báze – analogizuje v Jakobsonově chápání distinktivní rysy. Tuto roli v přirozených jazycích zastupují akustické vlastnosti hlásek.

Proti Jakobsonově analogii, ve které kombinatorika bází odpovídá distinktivním rysům, a kterou jsme v předchozí kapitole podrobili kvantitativnímu zkoumání, můžeme nalézt i další, zcela odlišné argu-

menty. Prvním z nich může být, že kombinatorní vlastnosti v DNA protilehlých bází jsou pouze konstrukčním principem DNA a nemají jinou než fyzikální opodstatněnost. Dvoušroubovice DNA je díky vazbě protilehlých bází odolná proti vlivům vnějšího prostředí a bezpečněji přenáší genetickou informaci. Oproti ní je jednovláknová RNA vůči vlivům prostředí mnohem náchylnější. Vztah protilehlých bází má funkci přidržovat obě vlákna DNA v dvoušroubovici, není to tedy funkce znaková. Identifikaci znaku a jeho složek zajišťuje komutační test, který jsme představili v první kapitole. Na základě jeho využití zjišťujeme, že o přiřazení aminokyseliny do proteinu a funkci proteinu rozhoduje kombinace bází v sekvenci DNA. K sobě vztažené pyrimidiny a puriny nemají žádný vztah k výslednému proteinu, neboť proteosyntézy se účastní fyzicky i sémioticky pouze jedno rameno DNA. Vztah pyrimidinů a purinů proto nemůžeme v gramatickém modelu DNA a proteosyntézy zohlednit. Do tabulky genetického kódu (viz instrumentarium v předchozí kapitole), která ukazuje tripletety kódující aminokyseliny, je vztah pyrimidinů a purinu nedosaditelný – nemá roli v kódu.

Dalším bodem Jakobsonovy analogie je průměr slov a tripletů. V DNA se vyskytuje 64 tripletů, které jsou realizovány jako trojkombinace čtyř bází. Triplet odpovídá jedné z dvaceti aminokyselin. Více tripletů tak kóduje jednu aminokyselinu, Jakobson tak do své analogie přidává koncept synonymie jakožto důsledek nekorespondence počtu tripletů a aminokyselin. Třem z tripletů Jakobson deleguje funkci delimitační/intonační, odpovídající ovšem funkci hlásky. Těmito delimitačními prostředky jsou odděleny syntaktické jednotky – věty odpovídající genům.

Přes Jakobsonovo problematické vršení analogií lze v relativní celistvosti rekonstruovat jeho gramatiku DNA a proteosyntézy. Činí tak např. Katz (2008). V následující části textu rekonstrukci této gramatiky představíme. Je zřejmé, že je taková gramatika v rozporu s některými sémiotickými principy, které jsme formulovali v předchozí části textu. Tato gramatika je tedy předem odmítnuta, odpovídá ale intuitivnímu a v nejrůznějších sémiotických přístupech k DNA nejčastěji se objevu-

jícímu pojetí. Její stručné představení můžeme využít k představení rozdílu přístupu k tvorbě gramatiky od výrazové a od obsahové roviny.

Výrazovou strukturou je v Jakobsonově gramatice jedno z ramen DNA. Obsahem je řetězec aminokyselin tvořící protein. V této gramatice platí jednoznačné přiřazení řetězce bází řetězci aminokyselin. Výraz i obsah zde mají odpovídající si hierarchičnost, která je motivována analogií jednotek národního jazyka – hláskami, slovem a větou. Výrazová rovina DNA a proteosyntézy je strukturována na báze, tripletu a genu. V obsahové rovině odpovídá bázi funkce odlišovat aminokyseliny (analogicky hlásky rozlišují slova). Tripletům odpovídají jednotlivé aminokyseliny (analogicky ke slovu). Kompletnímu řetězci aminokyselin proteinu odpovídá funkce proteinu (analogicky k větě). Z minimálního množství hlásek a bází se skládá větší množství slov a tripletů, které jsou dostupným slovníkem vyjadřování se za pomoci jednoho z kódů. Geny a věty mají komunikační autonomii, jsou to celistvé sémiotické reakce či exekuce kódu v prostředí.

Pro přehlednost předchozí zopakujeme. Báze má na obsahové rovině distinktivní funkci, její záměna mění aminokyselinu. Při transkripci je záměna báze příčinou dosazení jiné aminokyseliny do proteinového řetězce. Záměna cytosinu (C) za guanin (G) v tripletu GCU vede při translaci k umístění aminokyseliny glycin místo aminokyseliny alanin do proteinového řetězce. Podobně jako hlásky *e*, *o* a *a* odlišují slova *rek*, *rok* a *rak* odlišují báze G, C a U (uracil) aminokyseliny alanin (GCU), glycin (GGU) a valin (GUU).

Tripletu odpovídají přímo konkrétní aminokyseliny. Výraz tripletu má za obsah jednu z aminokyselin, která je dosazena do proteinového řetězce při translaci. Aminokyseliny mají přímý vliv na podobu prostorového uspořádání vlákna proteinu. Aminokyselina je tedy prostředek obsahu, který může být popsán svým potenciálem měnit prostorové uspořádání proteinu.

Gen má ve formě řetězce bází, respektive tripletů, obsah v podobě proteinu. Na základě svého prostorového uspořádání protein intera-

guje s dalšími fyzickými účastníky procesů v organismu. Obsahem genu je tedy funkce proteinu v organismu.

V tuto chvíli jsme představili základní gramatický model, který nám dále může sloužit jako ilustrace gramatického popisu DNA a proteosyntézy. Explicitně jsme uvedli, co v něm můžeme považovat za obsahovou a naopak za výrazovou rovinu. V následujícím výkladu se zaměříme na to, jakým způsobem je obecně možné tvořit gramatiku a jak se postup tvorby gramatiky vztahuje k DNA a proteosyntéze. Klíčovým problémem, kterým se budeme zabývat, jsou rozdíly v postupu tvorby gramatiky od výrazu, a naopak od obsahu.

Tvorba gramatiky

Označení gramatika tvořená od obsahu a gramatika tvořená od výrazu vyjadřuje, která ze sad je při tvorbě gramatiky preferována jako výchozí. Její strukturace je výchozí pro hledání funkcí mezi oběma sadami. Postup tvorby gramatiky pak předpokládá, že je struktura druhé sady analogická struktuře sady výchozí, anebo že je na základě využití struktury výchozí sady struktura druhé sady identifikovatelná.

Volba výchozí sady souvisí s její dostupností. V národních jazycích je z tohoto hlediska pro tvorbu gramatiky výhodnější výrazová sada. K výrazové sadě totiž máme zdánlivě přímý přístup. Výrazové veličiny počínaje hláskami a konče texty jsou zkušenostní daností, v jistém slova smyslu jsou samovolně identifikovány při učení se jazyku apod. Naopak obsahová rovina má hypotetický charakter. K jejímu popisu se dobíráme za pomoci aparátu, který může být libovolně konstruován. Testování identity veličin obsahové sady navíc není tak snadné jako je tomu u výrazu. Při testech se budou mluvčí častěji jednoduše vyjadřovat k odlišnosti hlásek *p*, *t*, *k*, než k odlišnosti obsahů např. takových slov jako *svižný*, *prudký*, *čilý*, *hbitý*, *rychlý*. Různá pojetí obsahu budou nabízet různé způsoby odlišení těchto slov. Nejistota o podobě strukturace obsahové sady posiluje tendenci vyjadřovat rozdíly mezi

předchozími slovy na základě strukturace jejich výrazové sady. Výše zmíněné hlásky *p*, *t*, a *k* jsou neznělé explozivy, rozdíl mezi nimi určuje jejich místo tvoření. Tím je jednoznačně vyjádřitelná jejich identita. I mluvčí bez teoretických znalostí fonetiky identifikuje jejich odlišnost. U zmíněných synonym bychom mohli tvrdit, že všechna mají význam dynamického pohybu a *hbitý* navíc příznak aktuálnosti a uzavřenosti popisovaného děje vyjádřeného verbálním substantivem, *svižný* by při záměně mělo význam neaktuální, trvalé dynamické potenciality subjektu pohybu, *čilý* by k významu pohybu přidávalo emocionální nádech a *prudký* časově blízké dosažení závěrové situace vyjadřované slovesem. Popis sémantických příznaků příkladových slov je analogický k vyjádření odlišnosti *p*, *t*, *k* na základě jejich místa tvoření. Je zřejmé, že popis sémantických diferencí příkladových slov má oproti popisu míst tvoření značně spekulativní charakter a můžeme si představit libovolné další způsoby vyjádření jejich odlišnosti. To je pro gramatický popis jednoznačně problematické.

Některé gramatické přístupy programově odmítají práci na gramatickém popisu s touto sadou, jejíž strukturace je hypotetická, tedy se sadou obsahu, a to proto, že neexistuje přímá evidence o povaze jejích prvků. Generativní gramatika má např. od svého založení modelační postupy založeny na modelu jedné sady (její vývoj však vykázal přesun k druhé sadě, viz Lakoff 1971; Fillmore 1968). Dovolíme si jistou hyperbolu, která zvýrazní důvody zaujetí negativního postoje k práci na gramatickém popisu od obsahové roviny.

V reprezentativní syntaxi češtiny (Grepl – Karlík 1998) upozorňují autoři na slovesa s více predikátory. Predikátorem se myslí ze sémantického i formálního hlediska rozdílná varianta užití slovesa. Příkladem takového slovesa je sloveso *vést*. V následujících příkladech ukážeme některé jeho predikátory:

„Benjamin vede Benedikta do šachového kroužku.“

„Benedikt vede v šachách nad Benjaminem.“

„Benedikt vede Benjaminův šachový kroužek.“

V první příkladové větě má predikátor *vést* význam, který můžeme deskribovat takto: subjekt A je se subjektem B v pohybu směrem k jejich destinaci a A určuje trasu pro B. Predikátor *vede* si zde vynucuje valenční pozici pro větně členské role subjektu a objektu, které obsazují právě *Benedikt* a *Benjamin*. V druhé příkladové větě je nutným obsazením valenční pozice jen *Benedikt*, predikátor více pozic k obsazení nevynucuje, ačkoliv v příkladové větě obsazeny jsou. Predikátor druhé příkladové věty lze popsat takto: subjekt A je v interakci se subjektem B a z hlediska evaluace jednání je blíže k dosažení cíle. Třetí příkladová věta má sémantický popis podobný příkladové větě první, určení trasy se zde však proměňuje v určování vývoje jednání skupiny. Sloveso *vést* má tedy v abstrakci význam orientovaného pohybu či jednání subjektu a participantů jeho orientovaného pohybu či jednání. Proto sloveso *vést* hodnotíme jako jednu lexikální jednotku. Všechny její predikátory participují na společném sémantickém jádře a představují jeho konkréti-zaci.

Navážeme na předchozí odstavce a zkonstruujeme sémantické popisy tak, aby vznikla pochybnost o určení výrazů konkrétních znaků. Lexémy *lak*, *vak* a *tak* mají zdánlivě zcela jiný obsah. Slovo *lak* vyjadřuje prostředek k nanášení na objekty pro jejich ochranu a designový benefit. Význam slova *vak* lze vystihnout jako předmět k uchovávání i přenášení jiných předmětů. Částice *tak* je navazovací prostředek, synonymní s *tedy*, který vyjadřuje následnost informačního obsahu výpovědi za předcházející. Pro *lak*, *vak* a *tak* můžeme podobně jako u lexému *vést* zkonstruovat abstraktní a všechny tři lexémy zahrnující sémantický popis. *Lak*, *vak* a *tak* vyjadřují prostředek udržení jistých objektů v kompaktní struktuře. *Vak* je prostředek pro uzavření k tomu vhodných předmětů ve svém nitru. *Lak* je prostředek, který pokrývá povrch předmětu, tyto předměty jsou lakem obemknuty, čímž se vytváří krycí a ochranný efekt. Významový popis částice *tak* uvedeme následujícím příkladem:

„Neměli jsme žádné rozkazy, velení na naši misi zcela zapomnělo a věděli jsme, že se dnes podává chutný oběd. Tak jsme se rozhodli vrátit se na základnu.“

Částice *tak* v tomto příkladě zahrnuje všechny obsahy předchozího sdělení. Představuje formální prostředek ohraničení všech obsahů předchozích výpovědí a plní tak roli anaforického substitutu. *Tak* má tedy stejné sémantické jádro jako *lak* a *vak*, je to prostředek udržení jistých objektů v kompaktní struktuře.

Po těchto sémantických popisech následují otázky ke gramatickému popisu. Tyto otázky se vztahují k tomu, jak identifikovat jazykové jednotky, pokud nemáme jistotu o sémantických rozdílech jednotlivých slov.

Pokud by slova *lak*, *vak* a *tak* byla sémanticky totožná, jak ukázala naše abstraktní deskripce jejich významu, pak bychom při komutačním testu nemohli zjistit odlišnost fonémů *l*, *v* a *t*. A jsou-li slova *vak*, *lak* a *tak* samostatné lexémy, proč jimi nejsou také predikátory slovesa *vést*?

Naše argumentace je v jistém smyslu extrémní. Názorně však ukazuje nejistotu, která vede k odmítnutí modelování gramatického systému na základě obsahové roviny popisované různými prostředky. Tímto je podtrženo intuitivní tláhnutí gramatiků k přístupu ke gramatice od výrazu, který představuje bezpečnou výchozí pozici.

Nyní se pokusíme na konkrétním gramatickém příkladu předvést, v čem spočívá problematičnost přístupu ke gramatickému popisu z výrazové roviny. Volíme si pro to gramatickou kategorii pádu, která je navíc prominentním příkladem z dějin gramatických popisů, co se rozmanitosti přístupu k ní týče. Její konceptuální i dějinný vývoj je strukturován právě opozicí přístupu od výrazu nebo od obsahu.

Vezmeme-li třídu českých substantiv, zjišťujeme, že se u nich pravidelně proměňuje zakončení. Pozorujeme-li souvislost zakončení substantiv s jejich obsahem, zjistíme, že odpovídá způsobu jejich zapojení ve větě. Systém zakončení substantiv je tvořený z hlásek či morfémů a vyvstává jako pádová struktura češtiny. V tomto případě je výraz výchozím modelačním východiskem pádového systému. Nemusíme dokonce specifikovat obsahy jednotlivých pádů, lze je nahrazovat

formálním zapojením substantiv do věty pojaté jako výrazový komplex. Takový popis pádu na základě funkce substantiva ve větě a za pomoci zakončení substantiv nalezneme například v *Příruční mluvnici češtiny* (Karlík – Nekula – Rusínová 2012, s. 235).

Zkoumáme-li výrazové proměny substantiv v různých pádových výskytech, ukazuje se, že se u nich neproměňuje pouze zakončení. Substantivum *pes* má i tvar *psem*, ve kterém je pádové zakončení vyjádřeno hláskami *e* a *m*, nebo můžeme obě hlásky zahrnout v morfému *-em*. Společně s tím se ale tvar substantiva mění tak, že je eliminováno *e* v kořeni. Tato změna je faktorem, který se podílí na vyjádření pádové funkce substantiva. K tomu se přidružují předložky, které opět mohou pádovou funkci spoluvyjadřovat. Např. *jezdít autem* má podobný význam jako *jezdít s autem* a jiný než *jezdít pod autem*. Pádový tvar ale ve všech třech příkladech odpovídá jedné pádové třídě – instrumentálu. Popisovat pádový systém na základě vztahu pádové funkce a zakončení substantiva se proto začíná jevit jako problematické. Z tohoto důvodu mnohé gramatické popisy pádu češtiny tyto výrazové proměny substantiv do pádového systému češtiny započítávají (Karlík – Nekula – Pleskalová 2002). Následující příkladové věty navíc ukazují, že se pád mění nejen na základě zakončení substantiva, změn kořene a užití předložky, ale děje se tak v souvislosti s celkovou výstavbou struktury věty (příklady jsou převzaty z práce Běličové 1982, s. 37).

„Horníci obsadili doly.“

„Horníci obsadili celé kupé.“

První příkladová věta vyjadřuje podřízení jisté lokace jednání určitých subjektů. Druhá příkladová věta vyjadřuje naopak lokální vyplněnost prostoru určitými subjekty. Tento popis obsahu obou příkladových vět ilustruje změnu pádové funkce. Na obsahové rovině nalézáme distinkci, která není dána zakončením, změnou kořene ani předložkou. Tato distinkce vyplývá ze vztahu substantiva k jinému substantivu. Pokud bychom se spoléhali na strukturaci pádového systému

pouze vzhledem k proměnám výrazových prostředků substantiva, nemohli bychom tuto distinkci identifikovat. Závěr je tady takový, že gramatiky konstruované od výrazu k obsahu mají tendenci nacházet jen takové obsahy, které jsou v přímém vztahu s nějakým výrazem. To je nejzávažnější argument proti gramatikám konstruovaným od výrazu – mají tendenci zanedbávat funkce obsahu s komplexní výrazovou strukturou výpovědi a delegují ji na bezprostředně nalézáný výraz.

Tvorba gramatiky DNA od výrazu a od obsahu

Nyní se vraťme zpět ke gramatice DNA a proteosyntézy. V této gramatice je výchozí výraz – řetězec DNA, strukturovaný na báze, triplety a geny. Tato strukturace určuje podobu obsahové sady. Báze má rozlišovací funkci písmene, triplet má funkci k aminokyselině a gen má funkci k proteinu. Obsahovou rovinu bychom pak v gramatice mohli popsat např. takto: báze má rozlišovací funkci v tripletu a obsah jí nepřísluší; obsahem tripletu je aminokyselina se schopností zapojovat se do řetězce s dalšími aminokyselinami; obsahem proteinu je jeho tvar. Takové pojetí gramatik DNA a proteosyntézy je zcela běžné v oblasti biosémiotiky či molekulární biologie, explicitně to vyjadřují např. Crickovy texty odkazované výš (1962, 1964, 1966, 1967, 1968). Toto pojetí gramatiky dále vede k pojetí obsahové roviny jako řetězce.

Pojetí obsahu jakožto řetězce ve svých pracích rozšiřuje Trifonov (1988, 1989, 1999). Jakákoliv interakce řetězce s řetězcem je pro Trifonova samostatným znakovým projevem. Vztahy DNA a RNA, RNA a RNA, včetně všech jejích variant (mRNA, snRNA, hnRNA), RNA a proteinového řetězce jsou znakovými projevy. Jak výrazová, tak i obsahová sada proto pro Trifonova představují řetězec.

Podobně jako Trifonov nachází Barbieri při modelování obsahové sady mnoho kódů přítomných v buňce, a to při procesech transkripce, translace, ale i u signálních faktorů buňky atd. (2005, 2006, 2008). Trifonovo a Barbieriho pojetí zohledňuje přítomnost výrazu a obsahu i ar-

bitrárnost jejich vztahu. Obsah i výraz ale pojmají jako řetězec a modelují znakové jevy od výrazové roviny, jednoho vybraného řetězce, který svou strukturu přenáší na druhý řetězec, který je s ním v interakci.

Místo pojetí obsahu, kterým byl výše řetězec, se nabízí jiné pojetí, ve kterém je obsahem chápána funkce daného řetězce. Konformace aminokyselinového řetězce (neboli protein) má svou funkci v organismu – transportní, stavební, signalizační apod. Např. protein transferin má význam přepravy železa v metabolismu. Obsah není popsán jako řetězec proteinu případně jeho konformace uzpůsobená k navazování atomů železa, obsahem je ale funkce transportu. Obsahová rovina tak nabývá zcela jiné podoby. Funkci proteinu můžeme dále nahlížet v širším kontextu, to znamená, jakou funkci má tento protein pro celý organismus. Význam transferinu pro celý organismus lze popsat jako reakci tohoto organismu na toxicitu prostředí, která je způsobena zvýšeným výskytem atomů železa. Variabilita popisu obsahové sady semiózy živého tvaru je tedy značně rozsáhlá.

Vraťme se k pojetí gramatiky, ve které bázi přísluší funkce rozlišovat triplety, proti tripletu stojí aminokyselina a proti genu protein. Pokud budeme obsahem rozumět funkci proteinu, pak je jeho výrazem celý řetězec genu. Nemůžeme se dále ptát na to, které části tohoto řetězce se přímo účastní zajišťování funkce. Gramatický popis nám tedy nedává možnost vyjádřit, jak např. určité triplety ovlivňují funkci proteinu. To je analogické k pádovým gramatikám, ve kterých vyjadřuje pád pouze koncovka substantiva a žádné další výrazové změny se vyjadřování pádu neúčastní – zapomíná se na roli proměny kořene, předložky a větného kontextu. Analogie spočívá v tom, že u obou případů vidíme strukturaci výrazové roviny, a tu bez dalších otázek přímo přiřazujeme obsahové rovině, bez toho aniž bychom se ptali, zda tento obsah nemá vztah ke komplexnější výrazové struktuře.

Konstrukce gramatiky ze sémantické báze

V předchozím textu jsme na základě příkladových vět a jejich sémantických popisů ukázali, jak je problematické uchopování obsahového roviny, které se místyjevilo jako libovolné. Stavysvěta je vždy nutné popisovat pomocí konceptů, v jiném případě by takový popis musel být nahrazen nekonečnou množinou fyzikálních vlastností, respektive tou množinou vlastností, které vnímá subjekt sémiózy a tím, jak na ni reaguje. Proto nemůže být obsahová sada pojímána jako řetězec, ale vždy musí vycházet z funkce, kterou má pro subjekt. Přes složitou povahu obsahu se ale ukáže, že přístup k tvorbě gramatiky od obsahu k výrazu je možný.

Přístup k tvorbě gramatiky směrem od obsahu k výrazu preferuje Kořenský: „V úvahách o ontologickém charakteru sémantického východiska je třeba postupovat v zásadě takto: aktivní uživatel jazyka má určité vyjadřovací potřeby; a právě „sémantický jazyk“ formulující konstrukční východisko daného typu lingvistických modelů představuje základní předpoklad k uskutečnění vyjadřovacích potřeb mluvčího v podobě základních relačních modelů stavů světa, o němž mluvčí vypovídá. Jde tedy o soubor potenciálních relačních objektů intenzionální povahy, které jsou mluvčímu k dispozici, jestliže vstupuje do komunikačních vztahů v rámci daného jazykového společenství. Bez existence tohoto dynamického, ale normovaného souboru relačních modelů stavů světa, tj. bez ustálených pravidel jazykového zobrazování skutečnosti, je komunikace stejně nemožná jako bez dynamických, ale normovaných zásad užití výrazových prostředků jazyka. Nelze předpokládat, že při složitých znakových vztazích současných přirozených jazyků lze chápat jazykové významy, které nejsou bezprostředně vázány na výrazové objekty, pouze jako aktuálně referenční, denotativní významy bez nezbytného zprostředkování významy intenzionálními. A právě tyto potenciální významy intenzionální povahy jsou v podobě sémantického východiska modelovány a stávají se konstrukční základnou modelů daného typu. Považujeme tedy za nezbytné přisuzovat sémantické bázi a pojímům jí funkčně ekvivalentním nejen gnozeologický,

ale i ontologický statut.“ (citujeme bez poznámek, Kořenský 1984, s. 13-14)

Předchozí odstavec představuje podstatnou námitku proti nerozhodnutelnosti o strukturaci obsahové roviny. Ta měla být zásadním faktorem, který způsobí odmítnutí modelování sémiózy od obsahové roviny, i když jsou známy nevýhody opačného směru. Kořenského přístup se pokouší odmítnout námitku, že gramatika nemá být tvořena od obsahu kvůli jeho hypotetičnosti oproti bezprostřední povaze výrazů.

Konstrukce gramatiky ze sémantické báze (Kořenský 1984) předpokládá, že význam má relační povahu – nejde o objekt jako takový, ale o relaci mezi kvalitami světa, kterými je tento objekt definován (např. barva, tvar, váha apod.). Např. slova *rozdělit*, *rozpūlit*, *rozseknout* jsou založena na stejné relaci, kterou můžeme popsat jako *mít tvar* a *mít dva tvary*. Soubor takových relací pak vytváří systém. Tento systém nám napomáhá strukturovat obsahovou sadu (vytváří významy). Jako možná se nám tak začne jevit i konstrukce gramatiky DNA a proteo-syntézy od obsahové sady.

Kořenský (1984) v Konstrukci gramatiky ze sémantické báze pracuje s obsahovou rovinou vztahenou k výrazové a pragmatické rovině. Následující příkladové věty jsou převzaty z Konstrukce gramatiky ze sémantické báze, pouze jsme je doplnili o kontext.

„Podívejte se, ten hrad se zřítíl.“

„Podívejte se támhle, vidíte zřícený hrad?“

„Hele, zřícenina.“

Kořenský ukazuje, jak předchozí tři příklady vyjadřují stejný význam: *rozbořená historická fortifikační a reprezentativní stavba*. Tento význam je vyjádřen slovtvorně, nepredikačně a predikačně. To znamená, že výrazově jsou významy vyjádřeny a/ slovtvorným utvořením lexému, nebo b/ syntaktickým spojením dvou lexémů, nebo c/ větně. V prvním případě se jedná o relaci tvaru *zřícený* a slovtvorného

prostředku *-ina*, který vyjadřuje výsledek děje: *zavařenina*, *zkamenělina*, *splynulina*. Ve druhém případě je vyjádření významu založeno na syntaktickém spojení dvou lexémů. Lexém *hrad* je adjektivem *zřícený* sémanticky specifikován opět do podoby výsledku děje. Ve třetím případě je využita větná realizace, která znovu vyjadřuje výsledek děje v podobě historické fortifikační a reprezentativní stavby ve stavu rozboření. Jeden význam z obsahové roviny je tedy vyjadřován různými výrazy. Výraz sám o sobě není v tomto gramatickém přístupu v žádném vztahu s dalšími možnostmi výrazové realizace téhož obsahu. Pokud bychom ke slovu *zřícenina* přistupovali z výrazové roviny, nemohli bychom zjistit, že je v jiných případech vyjadřován mimo slovo – syntagmatem nebo větou. Máme-li tyto možnosti vyjádření významu, můžeme dále sledovat jeho vztahy k pragmatické a výrazové rovině.

„Podívejte se, ten hrad se nejpozději zítra určitě zřítí.“

„Kdybyste se podívali támhle, uviděli byste zřícený hrad?“

„Ta ruina musí spadnout.“

První z příkladů vyjadřuje oproti jeho předcházející podobě změnu gramatické kategorie času. Druhý příklad je rozšířen o kondicionál, který nese aspekt podmíněnosti realizace děje predikátu výpovědi. Třetí příklad přesouvá realizaci významu ze slovotvorných prostředků, používá výraz pro zchátralou stavbu a proces zřícení lexikálně vyjadřuje jako spadnutí. Sémanticky je tak původní obsah plně realizován. První dva příklady mění pouze ztvárnění výpovědi ve smyslu vztahů jejich mluvčích, a to časových a modalitních. Třetí příklad mění lexikální základ zříceniny, zachovává ale vyjádřený obsah (doplněna je modalitu tvaru *muset*). Všechny tyto změny vyjádření *zříceniny* nemění její význam, ale její pragmatické zasazení do komunikace. Vyjádření významu i gramatiky zajišťuje výrazová rovina svými morfologickými tvary, syntaxí, a dalšími způsoby realizace textu.

Z hlediska přístupu k tvorbě gramatiky DNA a proteosyntézy od obsahu k výrazu se podobně jako v jazykovém případě nabízí vycházet od množiny jednoduchých stavů světa. Ty můžeme popisovat jako sou-

bor základních funkcí organismu či způsobů jeho chování. Takovou funkcí může být pohyb organismu, transport látek, přenos signálu, výstavba organismu, reakce na vnější stimul atd.

Organismus vybírá pouze některé vlastnosti světa, na které funkčně reaguje. Tyto vlastnosti rozlišuje na základě souboru základních funkcí. Např. transfer olova a transfer kadmia jsou založeny na stejné funkci, kterou je detoxikace. Podobně jako jsou slova *rozdělit*, *rozpůlit*, *rozseknout* založena na základní funkci *mít dva tvary*. Soubor takových základních funkcí není nutné vypisovat celý, ani Kořenský (1984) se o to v Konstrukci gramatiky ze sémantické báze nepokouší. Vzhledem k různým konfrontujícím se přístupům ke strukturaci sémantické roviny může být totiž souborů základních funkcí stanoveno nespočetné množství. Máme-li stanoveny funkce, které v organismu registrujeme, můžeme je považovat za obsah. K uzavření znakové funkce nám už chybí jen nalezení výrazové roviny. Za výraz můžeme považovat všechny procesy, které zajišťují naplňování funkce.

Pokud se vrátíme k původní představě gramatiky DNA, kde jsou výrazem báze, triplety a geny, musíme takto definovaný výraz rozšířit. V původním pohledu byl výrazem řetězec DNA. Díky jeho využití je naplňována určitá funkce.

Tento řetězec ale před svým využitím prochází řadou změn, jako je metylace nebo sestřih (tyto změny se navíc mohou v různých částech buňky lišit). Samotné využití řetězce je ovlivněno tzv. transkripčními faktory, které určují, zda bude řetězec využit k tvorbě proteinu s určitou funkcí, nebo ne. Funkce proteinu může být dále ovlivněna chaperonovými proteiny, které při konformaci proteinu ovlivňují jeho výsledný tvar. Díky tomuto vlivu na funkci proteinu budeme i tyto procesy považovat za součást výrazu. Výrazem je ve výsledku konečný tvar proteinu včetně všech procesů, které se podílely na zajištění jeho funkce. Tímto způsobem jsme na DNA a proteosyntézu aplikovali přístup ke gramatice od obsahu k výrazu. Původní představa genetického kódu v tom, co je jeho výrazem a obsahem, se tak podstatně proměnila. Za

obsah můžeme dál považovat funkci proteinu a za výraz prostředky, které ji zajišťují. Oprošťujeme se tak od pojetí znakového procesu v buňce pouze v souvislosti se zápisem v DNA a obracíme pozornost k užití tohoto zápisu ve funkcích, které jsou realizovány v živém tvaru.

Závěr

V celé knize jsme se snažili ukázat různé představy o tom, co můžeme považovat za gramatiku DNA a řešili jsme, co a z jakých důvodů můžeme považovat za genetický kód. Podívali jsme se na různé přístupy k tvorbě gramatik a zajímali jsme se o to, jak se dají použít k popisu DNA a procesů v živém organismu. S postupem výkladu jsme se stále více zaměřovali na to, co bychom měli považovat za obsah a za výraz genetického kódu. Zjistili jsme, že původní představy o genetickém kódu mají mnohá úskalí. Ukázalo se především to, že pohled na výraz i obsah genetického kódu jen jako na řetězec písmen (bází DNA a aminokyselin proteinů) je příliš zjednodušující, protože zcela paradoxně zapomíná na to, že tyto řetězce mají v organismu nějaké funkce. Ukázali jsme, že i když jsou k sobě v buňce tyto řetězceází a aminokyselin arbitrárně vztažené, nemusí být za obsah považován jiný řetězec. Jako obsah je možné vnímat výslednou funkci tohoto řetězce v organismu, stejně jako mají slova v přirozeném jazyce obsah ve svém použití. Dokonce ani výraz nemůžeme nakonec chápat jen jako řetězec, protože ten je v organismu používán za různých okolností a různým způsobem (např. exprese, sestřih). Všechny takové faktory ovlivňující jeho využití a podobu musíme započítávat do výrazu znaku.

Při popisu genetického kódu se původně vycházelo z vlastností, které jej činily kódem. Takovou vlastností byla arbitrárnost, linearita atd. Samotný kód se pak přiřadil tam, kde se tyto vlastnosti projevovaly, a to k lineárním řetězcůmází DNA a jejich arbitrárnímu vztahu k řetězci aminokyselin. Tím bylo pojetí genetického kódu uzavřeno. Pro kód je ale nejdůležitější vyjasnit, co je jeho obsahem a jaké výrazy tento obsah reprezentují. A pro to nejsou řetězceází ani aminokyselin dostatečné. Obsahem organismu je jeho chování a funkce, které naplňuje, např. přenášení kyslíku, za účelem dýchání. Z toho bychom měli při popisu genetického kódu vycházet a až následně bychom se měli ptát, jaké procesy v organismu tuto funkci umožňují realizovat. Pak budeme znaky v organismu popisovat jako funkcionální veličiny. Nové gramatiky DNA tak z tohoto pohledu už nemohou být jednodu-

chou tabulkou bází, tripletů a aminokyselin, ale budou komplexní sítě mezi výrazy, jejich expresí a funkcemi, způsobem jak k celému problému z praktických důvodů přistupuje současná bioinformatika.

Literatura

- Alberts, B. – Johnson, A. – Lewis, J. – Raff, M. – Keith, R. – Walter, P. (2008): *Molecular Biology of the Cell*. New York, NY: Garland Science.
- Barbieri, M. (2002): *The Organic Codes: An Introduction to Semantic Biology*. Cambridge: Cambridge University Press.
- Barbieri, M. (2005): Life is artifact-making. *Journal of Biosemiotics*, 1, s. 107–134.
- Barbieri, M. (2006): *Organické kódy*. Praha: Academia.
- Barbieri, M. (2008): Life is Semiosis. The biosemiotic view of Nature. *Cosmos and History. The Journal of Natural and Social Philosophy*, 4(1–2), s. 29–51.
- Běličová, H. (1982): *Sémantická struktura věty a kategorie pádu*. Praha: Academia.
- Berezovsky I., N. – Kirzhner, A. – Kirzhner, V., M. – Rosenfeld, V., R. – Trifonov E., N. (2003): Protein sequences yield a proteomic code. *J. Journal of biomolecular structure and dynamics*, 21(3): s. 317–25.
- Berg, J., M. – Tymoczko, J., L. – Stryer, L. (2012): *Biochemistry*. New York, NY: W. H. Freeman.
- Bolshoy, A. – Volkovich, Z. (V.) – Kirzhner, V. – Barzily, Z. (2010): *Genome Clustering from Linguistic Models to Classification of Genetic Texts*. Berlin – Heidelberg: Springer.
- Claverie, J., M. – Bouqueleret, L. (1986): Heuristic informational analysis of sequences. *Nucleic acids research*, 14(1), s. 179–96.
- Collado-Vides, J. (1993): A linguistic representation of the regulation of

- transcription initiation. I. An ordered array of complex symbols with distinctive features. *BioSystems*, 29, s. 87–104.
- Collado-Vides, J. (1992): Grammatical model of the regulation of gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 89(20), s. 9405–9409.
- Crick, F., H., C. (1966): Codon-anticodon pairing: the wobble hypothesis. *Journal of Theoretical Biology*, 19, s. 548–555.
- Crick, F., H., C. (1968): The Origin of the Genetic Code. *Journal of Theoretical Biology*, 38, s. 367–379.
- Crick, F., H., C. (1962): Towards the genetic code. *Discovery*, 22, s. 8–16.
- Crick, F., H., C. (1964): On the genetic code: Nobel lecture, December 11, 1962. In: *Nobel Lectures: Physiology or Medicine: 1942–1962*. Singapore – New Jersey, NJ – London – Hongkong: World Scientific, s. 811–21.
- Crick, F., H., C. (1967): The Croonian Lecture, 1966: The genetic code. *Proceedings of the Royal Society of London: Biological Sciences*, 167 (1009), s. 331–347.
- Crick, F., H., C. (1968): The origin of the genetic code. *Journal of Molecular Biology*, 38, s. 367–379.
- Cvrčková, F. (2006): *Úvod do praktické bioinformatiky*. Praha: Academia.
- Eco, U. (2009): *Teorie sémiotiky*. Praha: Argo.
- Faltýnek, D. (2010): Význam, vzduch a vývoj. In: Markoš, A. (ed.): *Jazyková metafora živého*. Červený Kostelec: Pavel Mervart, s. 181–188.
- Faltýnek, D. (2012): *Sémiotické primitivy v konstrukci gramatik*: Te-

- stování gramatik jazyka a DNA. Olomouc: Univerzita Palackého v Olomouci.
- Ferrer-i-Cancho, R. – Elvevag, B. (2010): Random texts do not exhibit the real Zipf's law-like rank distribution. PLoS ONE [online], 5(3), e9411. Cit. 8. 1. 2013. Dostupné z WWW: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009411>>.
- Ferrer-i-Cancho, R. – McCowan, B. (2009): A law of word meaning in dolphin whistle types. Entropy [online], 11(4), s. 688–701. Cit. 8. 1. 2013. Dostupné z WWW: <<http://www.mdpi.com/1099-4300/11/4/688>>.
- Ferrer-i-Cancho, R. (2006): When language breaks into pieces: a conflict between communication through isolated signals and language. BioSystems, 84(3), s. 242–253.
- Fillmore, C. (1968): The Case for Case. In: E. Bach – R. T. Harms (eds.): Universals in Linguistic Theory. New York: Holt, Rinehart & Winston, s. 1–88.
- Grepl, M. – Karlík, P. (1998): Skladba češtiny. Olomouc: Votobia.
- Grygar, M. (ed.) (1969): Pařížské rozhovory o strukturalismu. Praha: Svoboda.
- Hála, B. (1956): Slabika – její podstata a vývoj. Praha: Československá akademie věd.
- Hartl, D., L. – Ruvolo, M. (2013): Genetics: Analysis of Genes and Genomes [8. vydání]. Burlington, MA: Jones and Bartlett Learning.
- Havlin, S. – Buldyrev, S., V. – Goldberger, A., L. – Mantegna, R., N. – Peng, C-K. – Simons, M. – Stanley, H., E. (1995): Statistical and linguistic features of DNA sequences. Fractals, 3(2), s. 269–284.
- Hjelmslev, L. (1972): O základech teorie jazyka. Praha: Academia.

- Hoffmeyer, J. (2008): *Biosemiotics: An Examination into the Signs of Life and the Life of Signs*. Scranton: University of Scranton Press.
- Chomsky, N. (1966): *Syntaktické struktury*. Praha: Academia.
- Jakobson, R. (1973): *La linguistique et les sciences naturelles. Essais de linguistique générale*, Paris: Éd. de Minuit.
- Jakobson, R. (1971): *Linguistics in relation to other sciences*. In: Roman Jakobson, *Selected Writings: Vol. 2: Word and Language*. The Hague – Paris: Mouton, s. 655–696.
- Ji, S. (2000): *Free energy and information contents of Conformons in proteins and DNA*. *BioSystems*, 54, s. 107–130.
- Ji, S. (2006): *The Proteome as an Autonomous Molecular Language: „Proteinese“*. A Poster accepted for presentation at the DIMACS Workshop on Sequences, Structure and Systems Approaches to Predict Protein Function, Rutgers University, Piscataway, May 3–5.
- Ji, S. (1997): *Isomorphism between cell and human languages: molecular biological, bioinformatic and linguistic implications*. *BioSystems*, 44(1), s. 17–39.
- Ji, S. (1999): *The linguistics of DNA: words, sentences, grammar, phonetics, and semantics*. *Annals of the New York Academy of Science*, 870, s. 411–417.
- Karlík, P. – Nekula, M. – Pleskalová, J. (2002): *Encyklopedický slovník češtiny*. Praha: Nakladatelství Lidové noviny.
- Karlík, P. – Nekula, M. – Rusínová, Z. (2012): *Příruční mluvnice češtiny*. Praha: Lidové noviny.
- Katz, G. (2008): *The hypothesis of a genetic protolanguage: an epistemological investigation*. *Biosemiotics*, 1(1), s. 57–73.
- Konopka, A., K. (1995): *Noncoding DNA, Zipf's law, and language*. *Science*, 268(5212), s. 789.

- Kořenský, J. (1972): Komplexní popis výrazové struktury pádu substantiva v češtině. Praha: Academia.
- Kořenský, J. (1984): Konstrukce gramatiky ze sémantické báze. Praha: Academia.
- Kořenský, J. (1987): K procesuálnímu modelování řečové činnosti. Slovo a slovesnost, 48, s. 177–189.
- Kořenský, J. (1988): Procesuální modelování řečové činnosti a interpretace. Slavica Pragensia, 32, s. 125–135.
- Kořenský, J. (1992): K otázce procesuálního pojetí slovní zásoby. Slovo a slovesnost, 53, s. 265–272.
- Kořenský, J. (1994): Komunikativně orientovaný přístup ke gramatickým kategoriím. In: Gramatika a semantika na slavjanskite ezici v komunikativen aspekt. Sofia, s. 7–14.
- Kořenský, J. (1994): Morfologické kategorie a procesuálně-komunikační přístup k řeči a jazyku. Slovo a slovesnost, 55, s. 81–89.
- Kořenský, J. (1996): Pád pádu aneb kam s pádem v komunikativně orientované gramatice. Slovo a slovesnost, 57, s. 212–217.
- Kořenský, J. (1997): Kam se vlna obrací aneb nikoli anti-Beaugrande. Slovo a slovesnost, 58, s. 161–173.
- Kořenský, J. (1998): Proměny myšlení o řeči. Praha: FFUK.
- Kořenský, J. (2003). Procesuální gramatika v kontextu současných tendencí lingvistického myšlení. Slovo a slovesnost, 64, s. 1–7.
- Kořenský, J. (2006): Člověk – řeč – poznání. Olomouc: Univerzita Palackého.
- Kull, K. (2005). A brief history of biosemiotics. Journal of Biosemiotics, 1, s. 1–34.

- Lakoff, G. – Johnson, M. (2002): *Metafory, kterými žijeme*. Brno: Host.
- Lakoff, G. (1971): On generative semantics. In: Jakobovits, L. – Steinberg, D. (eds): *Semantics*. Cambridge: Cambridge University Press, s. 232–296.
- Lakoff, G. (2006): *Ženy, oheň a nebezpečné věci*. Praha: Triáda.
- Langacker, R., W. (2009): *Investigations in Cognitive Grammar*. Berlin: Mouton de Gruyter.
- Li, W. (1992): Random texts exhibit Zipf's-law-like word frequency distribution. *IEEE Transactions on Information Theory*, 38(6), s. 1842–1845.
- Mantegna, R., N. – Buldyrev, S., V. – Goldberger, A., L. – Havlin, S. – Peng, C-K. – Simons, M. – Stanley, H., E. (1994): Linguistic features of noncoding sequences. *Physical Review Letters*, 73(23), s. 3169–3172.
- Mantegna, R., N. – Buldyrev, S., V. – Goldberger, A., L. – Havlin, S. – Peng, C-K. – Simons, M. – Stanley, H., E. (1995): Systematic analysis of coding and noncoding DNA sequences using methods of statistical linguistics. *Physical Review: E, Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 52(3), s. 2939–2950.
- Markoš, A. – Grygar, F. – Hajnal, L. – Kleisner, K. – Kratochvíl, Z. – Neubauer, Z. (2009): *Life as its own designer: Darwin's Origin and Western thought*. Berlin: Springer.
- Markoš, A. – Švorcová, J. (2009): Recorded versus organic memory: interaction of two worlds as demonstrated by the chromatin dynamics. *Biosemitotics*, 2, s. 131–149.
- Markoš, A. (2003): *Tajemství hladiny*. Praha: Dokořán.
- Markoš, A. (2009): Do Biosemiotics, But Don't Forget Semiosis. *Biosemitotics*, 2, s. 117–119.

- Markoš, A. (ed.) (2008): *Náhoda a nutnost. Jacques Monod v zrcadle dnešní doby*. Červený Kostelec: Pavel Mervart.
- Markoš, A. – Faltýnek, D. (2011): Language metaphors of life. *Biosemiotics*, 4(2), s. 171–200.
- Markoš, A. (1997): *Povstání živého tvaru*. Praha: Vesmír.
- Markoš, A. (2002): *Readers of the Book of Life: Contextualizing Developmental Evolutionary Biology*. New York, NY: Oxford University Press.
- Maston, G., A. – Evans, S., K. – Green, M., R. (2006). Transcriptional Regulatory Elements in the Human Genome. *Annual Review of Genomics and Human Genetics* (7). Palo Alto: Annual Reviews, s. 20–59.
- Monod, J. (1970): *Le hasard et la nécessité: Essai sur la philosophie naturelle de la biologie moderne*. Paris: Éditions du Seuil.
- Niyogi, P. – Berwick, R., C. (1995): A note on Zipf's law, natural languages, and noncoding DNA regions [online]. A. I. Memo, (1530) / C.B.C.L. Paper, (118). Cit. 8. 1. 2013. Dostupné z WWW: <<ftp://publications.ai.mit.edu/ai-publications/pdf/AIM-1530.pdf>>.
- Nöth, W. (1990): *Handbook of Semiotics*. Bloomington: Indiana Univ. Press.
- Palazzo, A., F. – Gregory, R. (2014): The case for junk DNA. *PLoS Genetics*, 8, 10(5), e1004351.
- Pattee, H., H. (2001). The Physics of Symbols: Bridging the Epistemic Cut. *Biosystems*, 60, s. 5–21.
- Pattee, H., H. (2008): Physical and Functional Conditions for Symbols, Codes, and Languages. *Biosemiotics*, 1(2), s. 147–168.
- Popov, O., S. – Segal, D., M. – Trifonov, E., N. (1996): Linguistic com-

- plexity of protein sequences as compared to texts of human languages. *BioSystems*, 38(1), s. 65–74.
- Raible, W. (2001): Linguistics and genetics: systematic parallels. In: Hapspelmath, M. – König, E. – Oesterreicher, W. – Raible, W. (eds.): *Language Typology and Language Universals: An International Handbook / Sprachtypologie und sprachliche Universalien: Ein internationales Handbuch / La typologie des langues et les universaux linguistiques: Manuel international*. Berlin – New York: Walter De Gruyter, s. 103–123.
- Richard, P. – Manley, J. (2009): Transcription Termination by Nuclear RNA Polymerases. *Genes & Develoepment*, 23, s. 1247-69. Dostupné z WWW: <<http://www.genesdev.cshlp.org/content/23/11/1247.long>>.
- Searls, D., B. (1997): Linguistic approaches to biological sequences. *Computer applications in the biosciences*, 13, s. 333–344.
- Searls, D. B. (2001): Reading of the book of life. *Bioinformatics*, 17(7), s. 579–580.
- Searls, D., B. (2002): The language of genes. *Nature*, 420, s. 211–217.
- Shulman, M., J. – Steinberg, C., M. – Westmoreland, N. (1981): The coding function of nucleotide sequences can be discerned by statistical analysis. *Journal of Theoretical Biology*, 88, s. 409.
- Schrödinger, E. (1944): *What is Life?* Cambridge: Cambridge University Press.
- Stanford, A., L. (1975): *Foundations of Biophysics*. New York: Academic Press.
- Strachan, T. – Read, A. (2004): *Human Molecular Genetics 3*. London: Garland Press.
- Těšitelová, M. (1992): *Quantitative Linguistics: Linguistic and Literary*

- Studies in Eastern Europe. Amsterdam a Philadelphia: John Benjamins Publishing Company.
- The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, s. 57–74.
- The european bioinformatics institute (EMBL-EBI) (2014). Cit. 19. 10. 2014. Dostupné z WWW: <[http://www.ebi.ac.uk/ena/data/warehouse/search?query=tax_eq\(9606\)&domain=coding&result=coding_release&display=fasta&download=zip](http://www.ebi.ac.uk/ena/data/warehouse/search?query=tax_eq(9606)&domain=coding&result=coding_release&display=fasta&download=zip)>.
- Trifonov, E., N. (1989): The multiple codes of nukleotide sequences. *Bulletin of Mathematical Biology*, 51, s. 417–432.
- Trifonov, E., N. (1999): Elucidating sequence codes: three codes for evolution. *Annals of the New York Academy of Sciences*, 870, s. 330–338.
- Trifonov E., N. – Berezovsky I., N. (2002): Proteomic Code. *Molecular Biology*, 36, s. 239–243.
- Trifonov, E., N. (1988): Codes of nucleotide sequences. *Mathematical Biosciences*, 90 (1-2), s. 507–517.
- Tsonis, A., A. – Elsner, J., B. – Panagiotis, A. (1997): Is DNA a language? *Journal of Theoretical Biology*, 184 (1), s. 25–29.
- Twyman, R., M. (1998): *Advanced Molecular Biology: A Concise Reference*. New York, NY – Abingdon: Taylor and Francis.
- Twyman, R., M. (2004): *Principles of Proteomics*. Abingdon – New York, NY: Garland Science / BIOS Scientific Publishers.
- Uexküll, J. von (2006): Nauka o významu. In: Kliková, A. – Kleisner, K.: *Umwelt. Koncepce živého světa Jakoba von Uexküllä*. Červený Kostelec: Pavel Mervart, s. 13–69.
- Watson, J., D. – Berry, A. (2003): *DNA: The Secret of Life*. New York, NY: Alfred A. Knopf.

Weaver, R., F. (2002): *Molecular Biology*. Boston, MA: McGraw-Hill.

Zipf, G., K. (1949): *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press.

Resumé

Kniha předkládá návrhy ke změně pojetí genetického kódu. V první řadě se zaměřuje na způsob analogizování genetického textu a přirozeného jazyka. První kapitola rozebírá úlohu hlásky v jazykovém systému a složení hlásky z distinktivních rysů - včetně jejich proměny v závislosti na kombinatorice hlásek v textu a pozici ve slově. Snažíme se ukázat, že hláskový systém je vhodné popisovat včetně kombinatorických vlastností jednotlivých hlásek, protože jen tak můžeme hodnotit roli jednotlivých hlásek v systému. V souvislosti s tím ukazujeme úlohu hlásek naplňovat sémiotickou podmínku funkce kódu - arbitrárnost. Hláskový systém představuje implementaci této podmínky funkce kódu do jazykového systému, což ukazujeme na příkladu z ostrova Balnibarbi. Na konci první kapitoly představujeme analogii hlásky a bází DNA, případně aminokyselin. Na základě předcházející argumentace se snažíme rozhodnout, zda lze roli hlásky v genetickém kódu přisoudit hlásce, nebo aminokyselině.

V druhé kapitole vycházíme z pojetí genetického kódu navrženého F. Crickem, ukazujeme ale další pojetí kódu v živém organismu vybraných biologů, biosémiotiků nebo lingvistů. Pro základní přehled předkládáme čtenáři základní popis proteosyntézy. Zároveň se snažíme resumovat předešlé kvantitativní lingvistické analýzy genetického textu (komplexita - entropie, Zipfův zákon apod.). Hlavní částí textu je analýza projevů Zipfova zákona v textu přirozeného jazyka a v mRNA. Tato analýza potvrzuje náš předpoklad, že není vhodné analogizovat hlásky a báze - že báze nejsou písmena.

V poslední části textu vysvětlujeme, co to znamená provádět gramatický popis, ukazujeme také, že o gramatickém popisu můžeme mluvit ve vztahu k popisu libovolného textu, včetně textu genetického. Zároveň čtenáři předkládáme jednoduchou teorii znaku. Rozebíráme Jakobsonův gramatický popis DNA, který jsme používali v předchozí kapitole. Dále srovnáváme dva přístupy k tvorbě gramatiky - přístup od výrazu a přístup od obsahu. Na příkladu gramatického popisu pádu ilustrativně ukazujeme výhody a nevýhody obou přístupů. Zároveň se

ptáme na to, jak by byl přístup od výrazu a od obsahu použit v případě gramatického popisu genetického textu. V souvislosti s tím argumentujeme pro to, že v gramatickém ani znakové popisu nemůže být jako obsahová složka znaku brán řetězec biopolymerů (bází, aminokyselin) - to je totiž pojetí znaku často se vyskytující v biosémiotice. Nakonec upozorňujeme na to, že přístup k tvorbě gramatiky od obsahu je vhodný pro popis genetického textu, vycházíme při tom z konceptů Jana Kořenského, který formuloval principy tvorby gramatiky ze sémantické báze.

Summary

In this book we show some proposals of the genetic code design. We examine the analogy of the natural language and DNA. At first we focus on the notion of the speech sound. Speech sound is described as a combination of distinctive features. We show that speech sound system includes combinations of the distinctive features in the specific speech sound and changes of distinctive features of the speech sound in the text (by combination of the speech sounds or position in the word). The first chapter presets double articulation as a necessary condition for the natural language (by Barnibalbi example) and speech sound system as an implementation of double articulation in the code. At the end of the chapter we describe analogy of the speech sound and character of DNA (base) or character of protein string (amino acid). By this discussion we try to find levels of double articulation in the semiosis inside the cell.

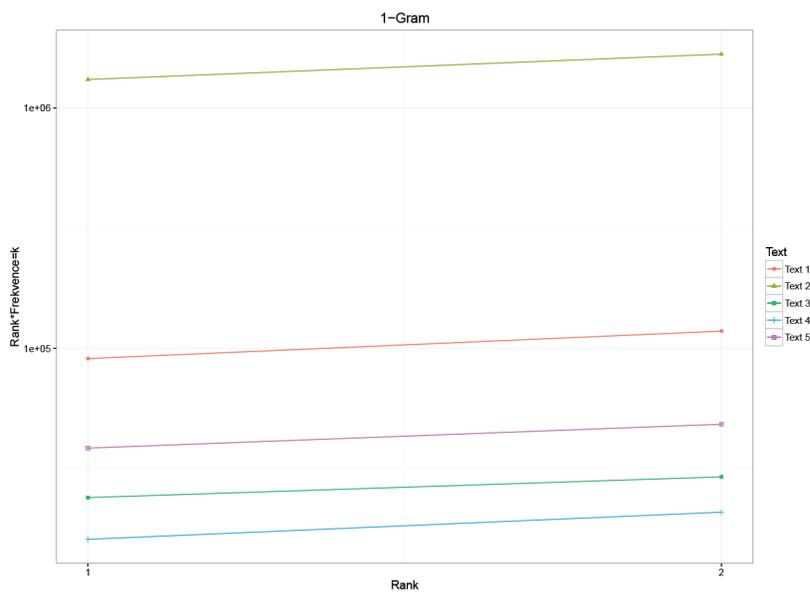
In the second chapter we proceed from the discovery of DNA structure to current stage of the molecular biology. Generally we introduce the basic semiotic assumptions of molecular biology in the description of the structure of DNA, proteins and genetic code. We focus on interpretations of F. Crick, another molecular biologists, biosemioticians and linguists. For the aims of the paper we describe some fundamentals of molecular biology. Core of our text is quantitative analysis (n-gram structure, Zipf's law) of mRNA strings and natural language text. We take into consideration representative quantitative analysis of DNA, RNA and proteins too. Our analysis of mRNA confirms the assumption that the design of the genetic code cannot analogize DNA bases and letters.

In the last chapter we explain why it is possible to consider any systematized description of the text, DNA and proteosynthesis as a grammar. For that reason we present the terms grammar and sign illustratively. We briefly introduce some grammatical conception of DNA and on an example of Jakobson's grammatical DNA model we put it to a semiotic analysis. The centre of the chapter is a confrontation

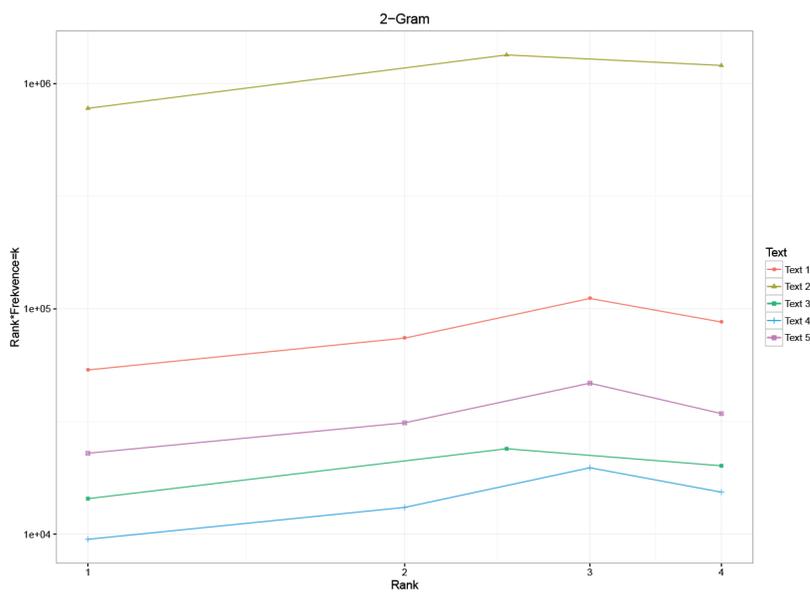
of approaches to grammar construction from the expression and from the content – by examples from the theory of the case we explain the advantages and disadvantages of both approaches and we transfer them onto the grammatical model of DNA and proteosynthesis. We state that the string (of bases, amino acids) cannot be considered as a content and expression level of the sign process of proteosynthesis. In the conclusion we determine the method of creating grammar from the content, i.e. from the function as the appropriate to grammatical description of DNA and proteosynthesis.

Přílohy

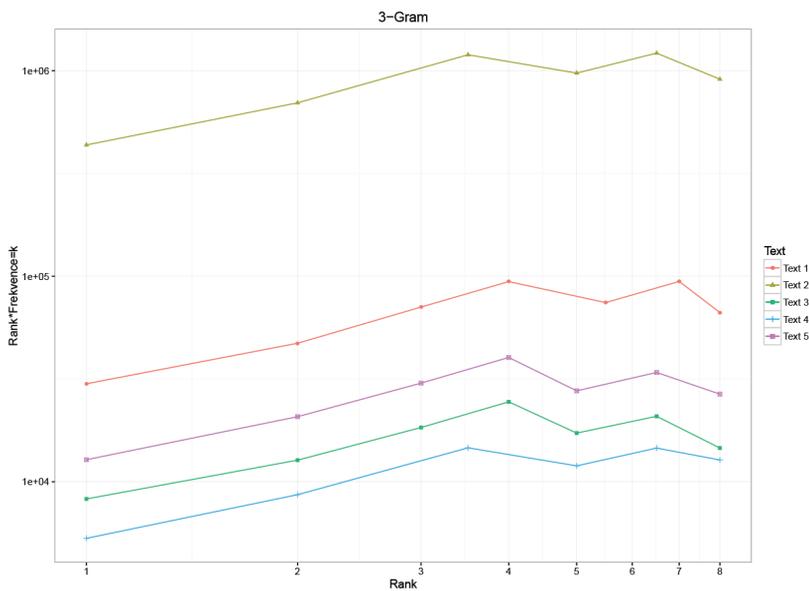
N-gramové analýzy textů přirozeného jazyka a mRNA



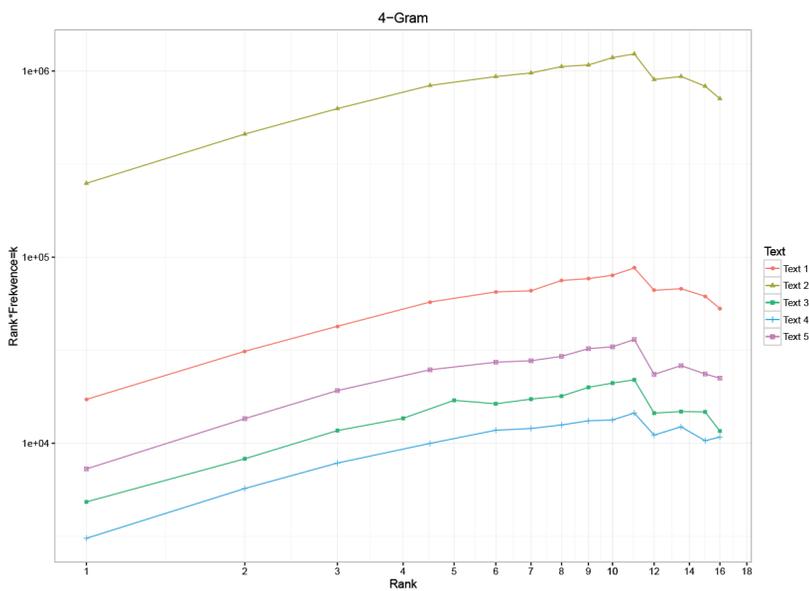
1-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



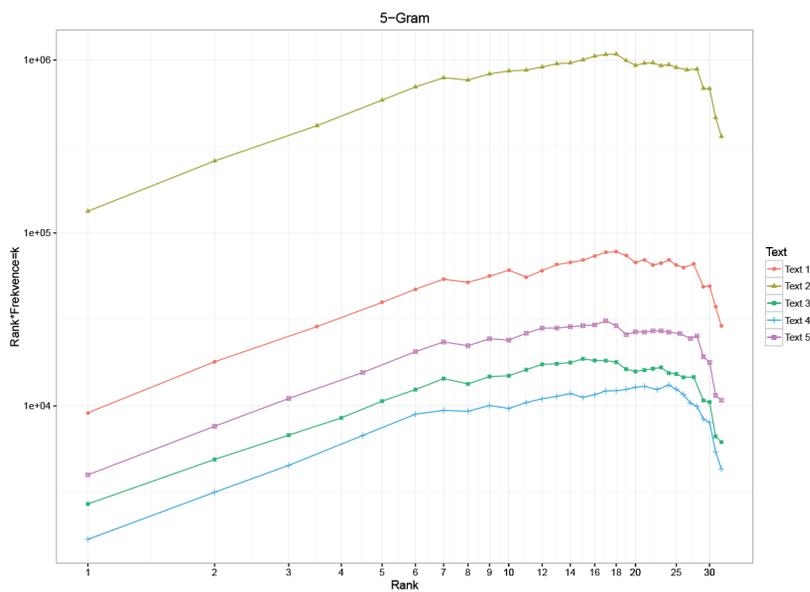
2-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



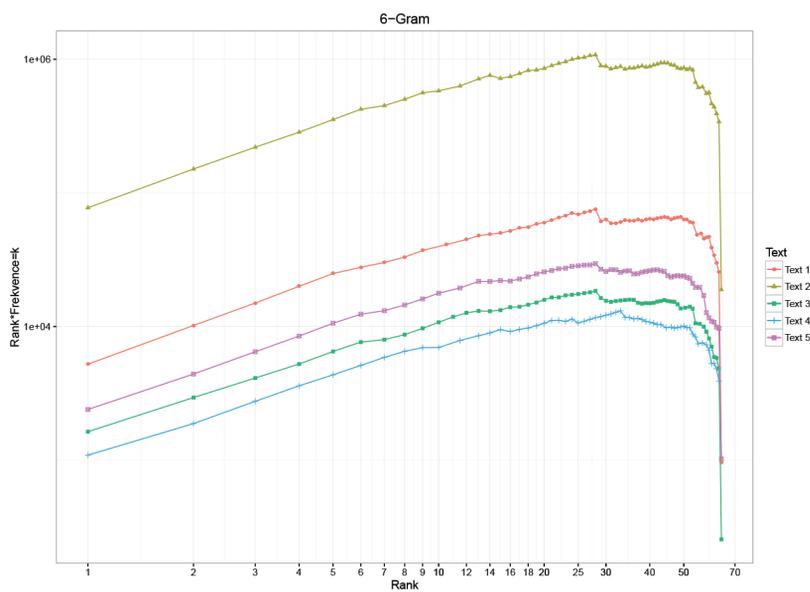
3-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



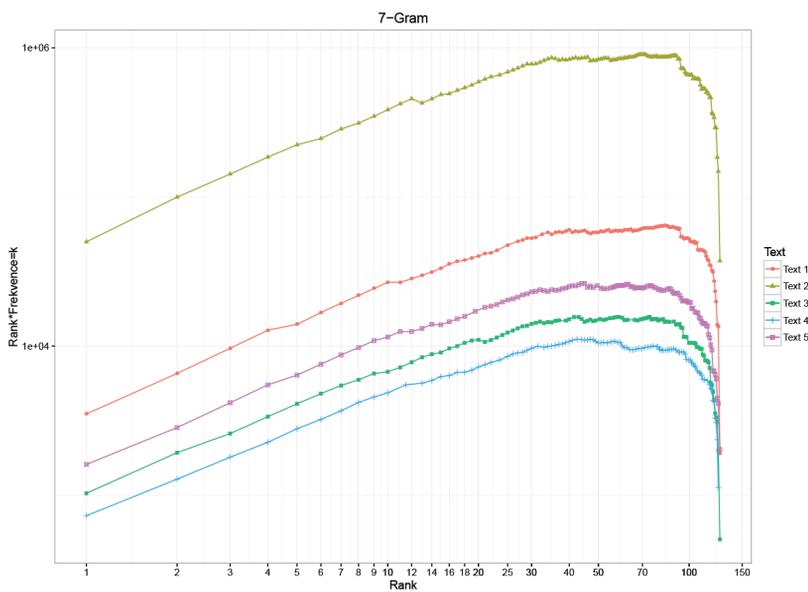
4-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



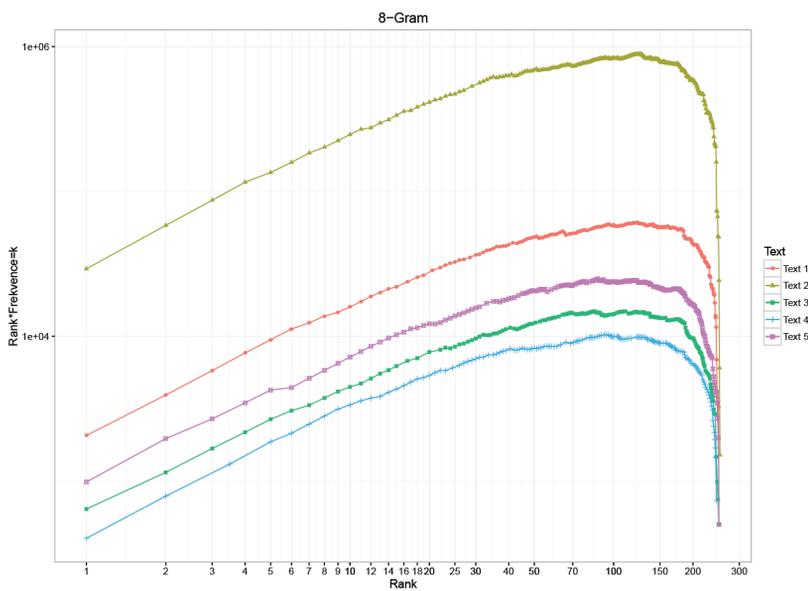
5-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



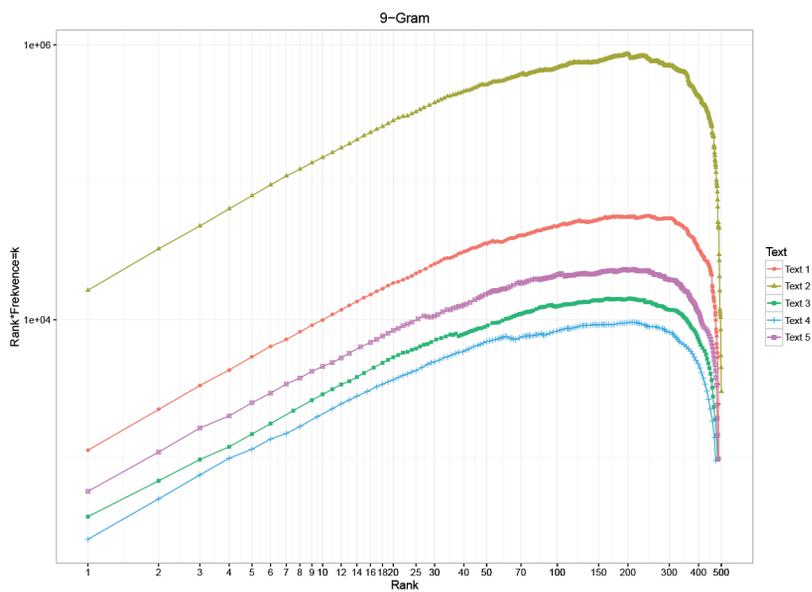
6-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



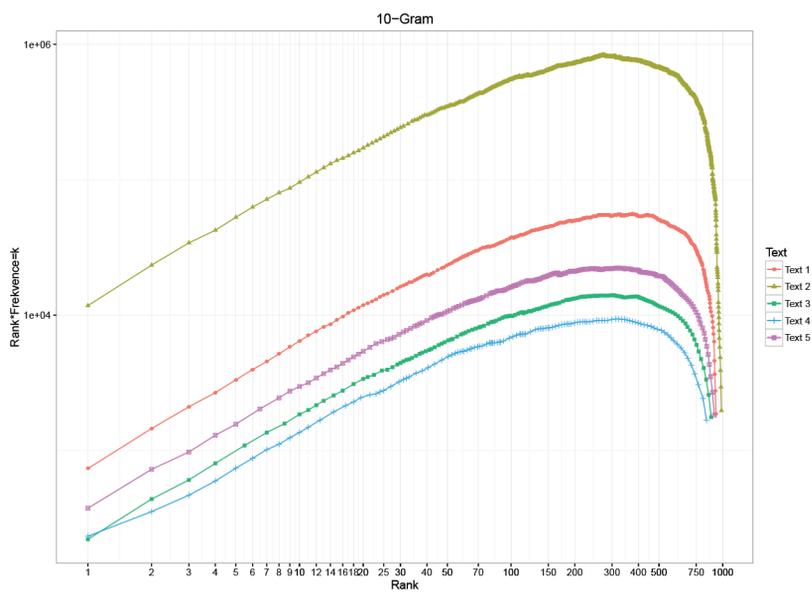
7-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



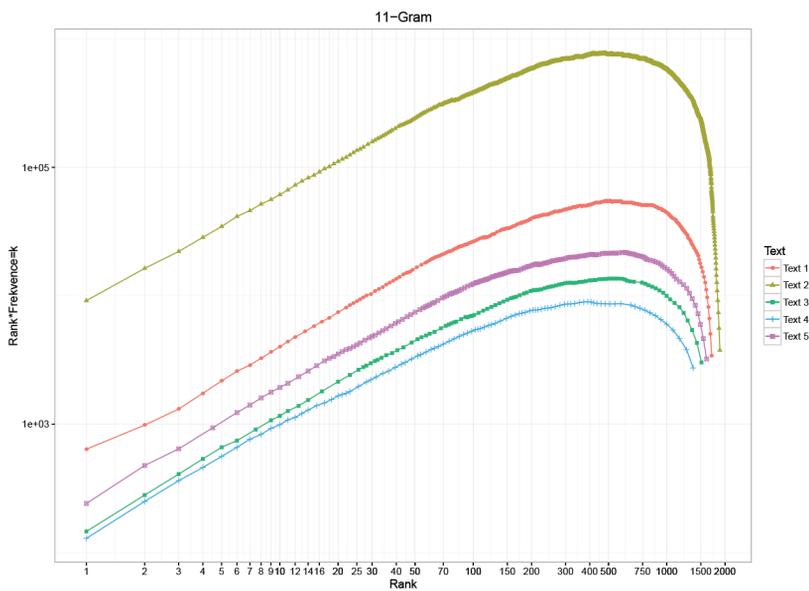
8-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



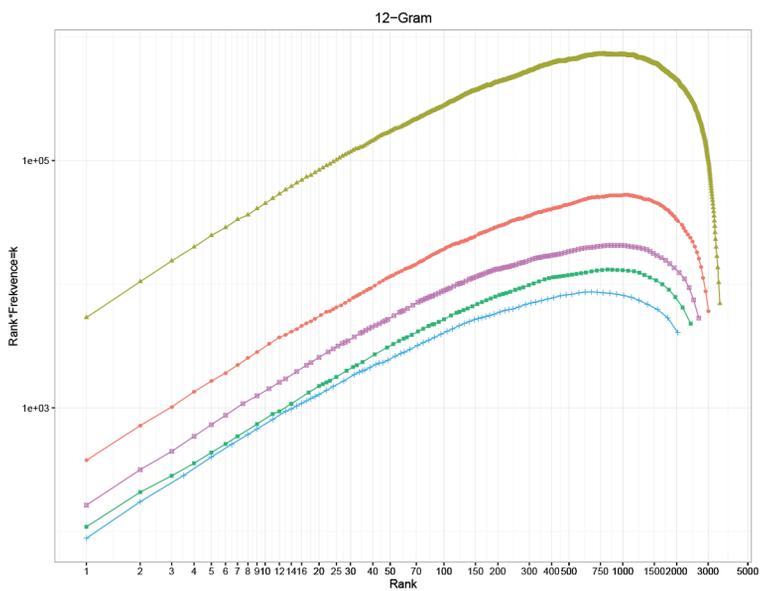
9-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



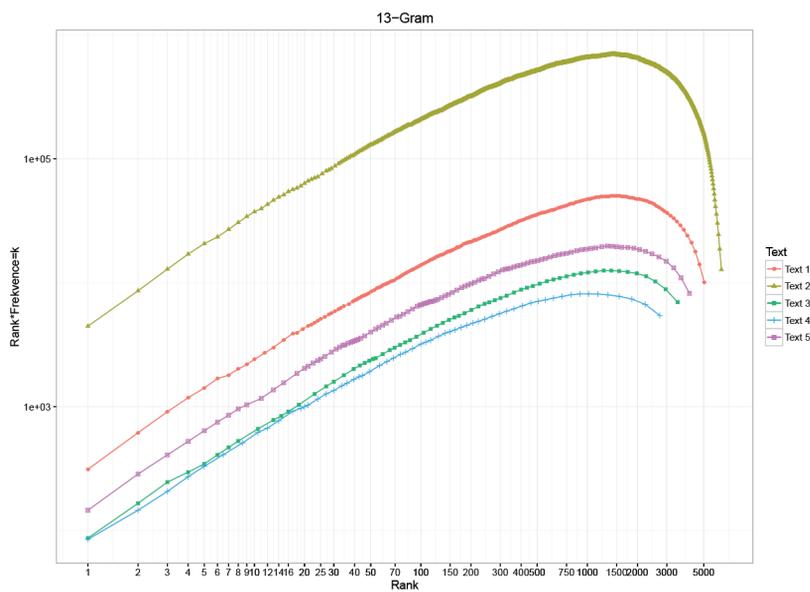
10-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



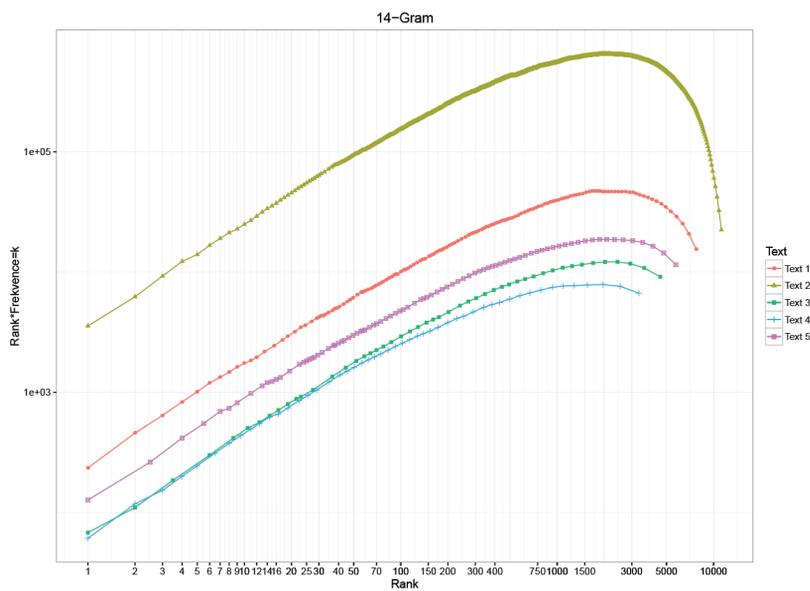
11-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



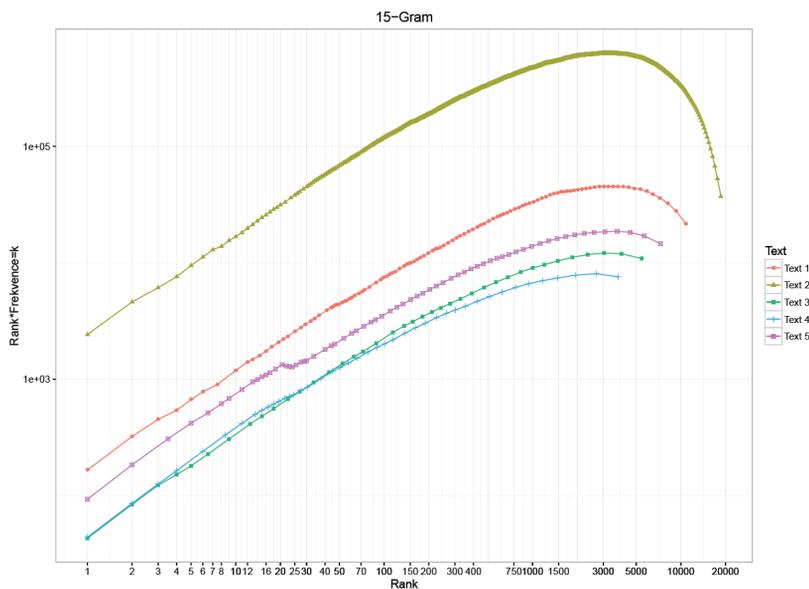
12-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



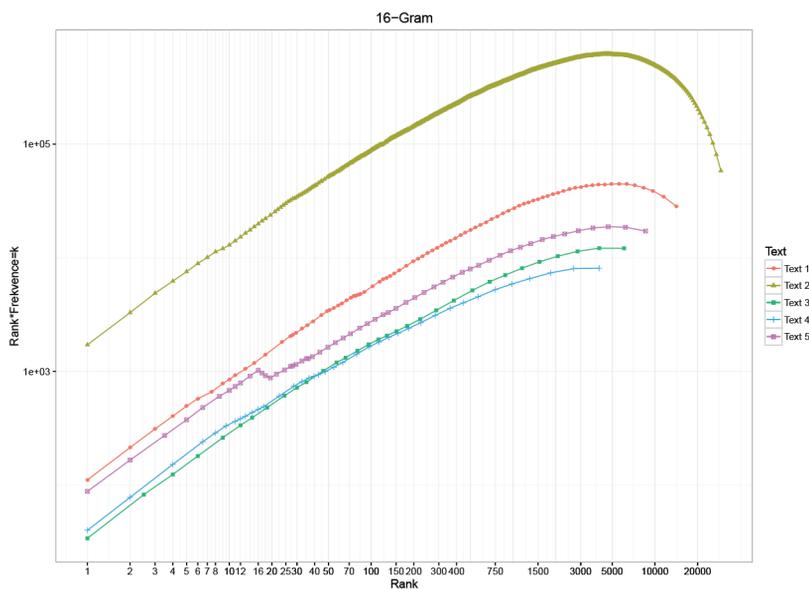
13-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



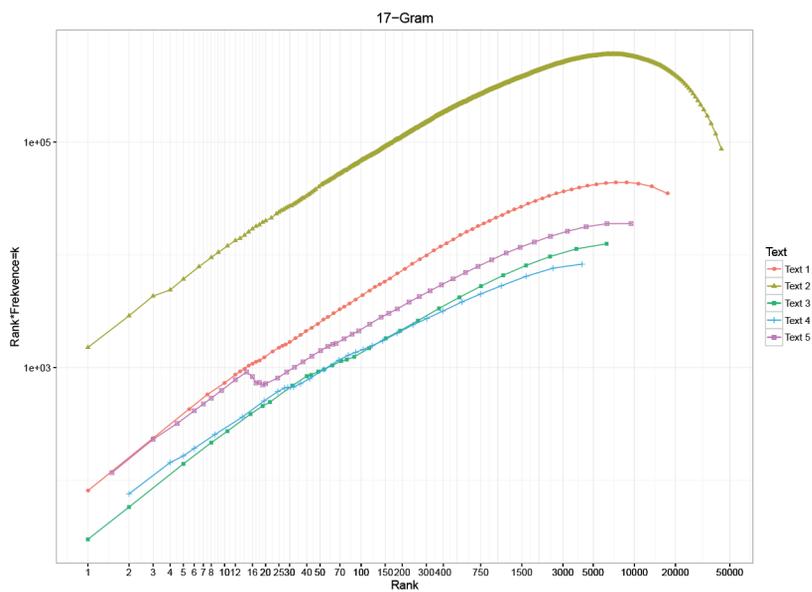
14-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



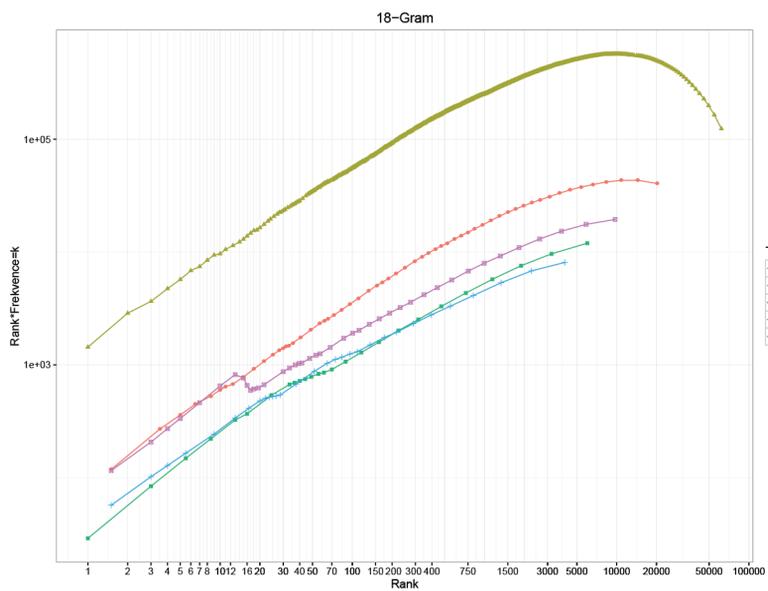
15-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



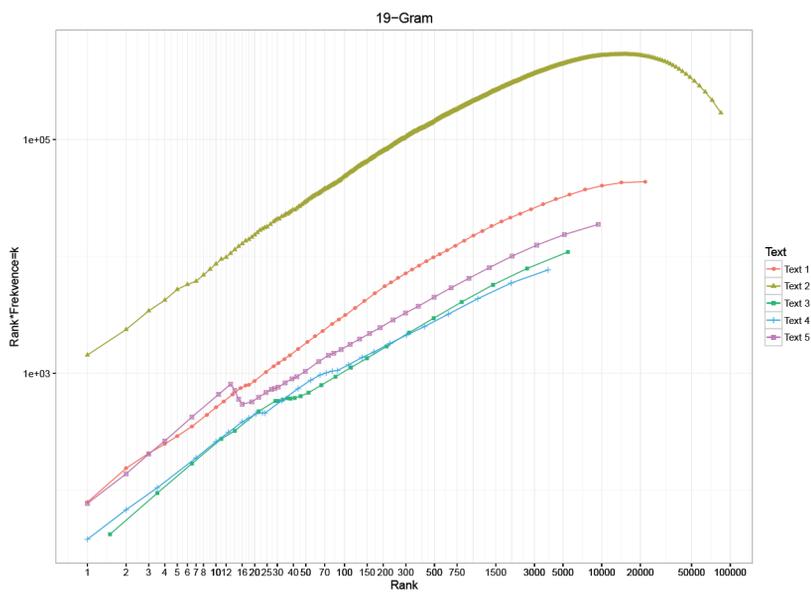
16-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



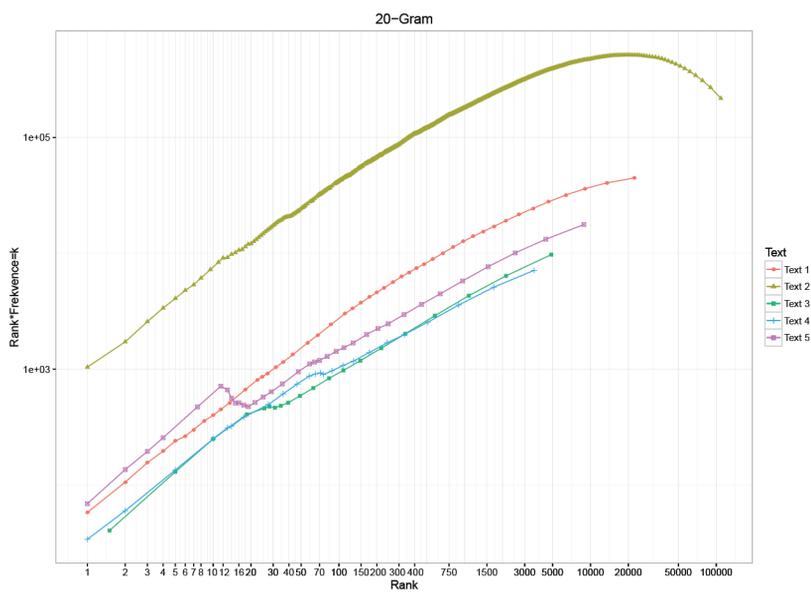
17-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



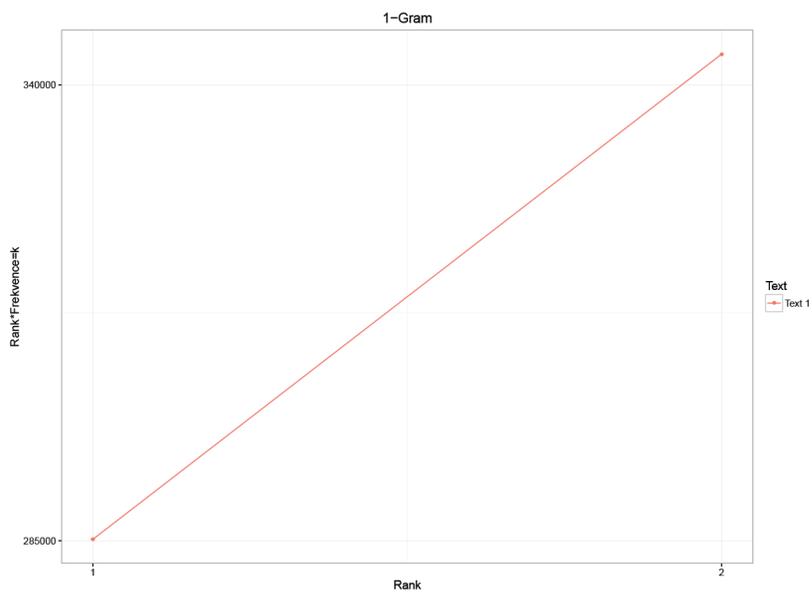
18-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



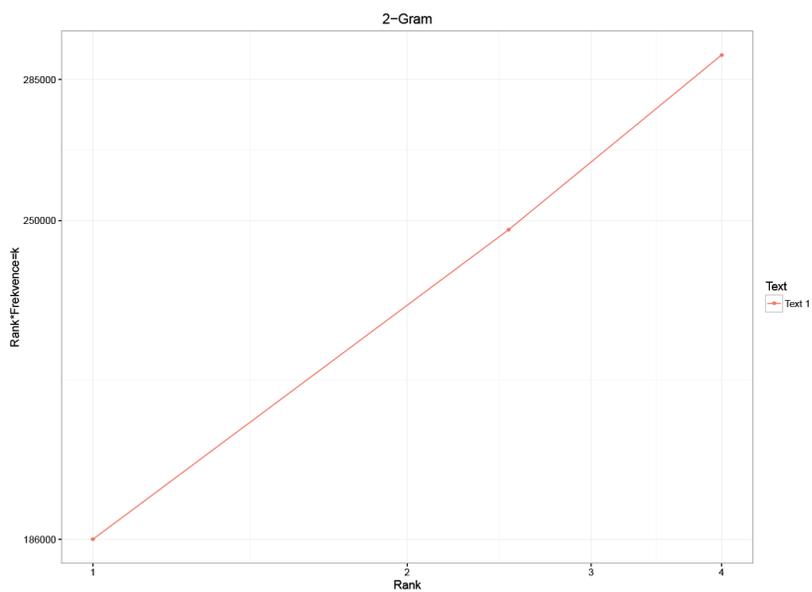
19-gram textů českého jazyka (texty byly náhodně vybrány z korpusu cca 1000 českých textů)



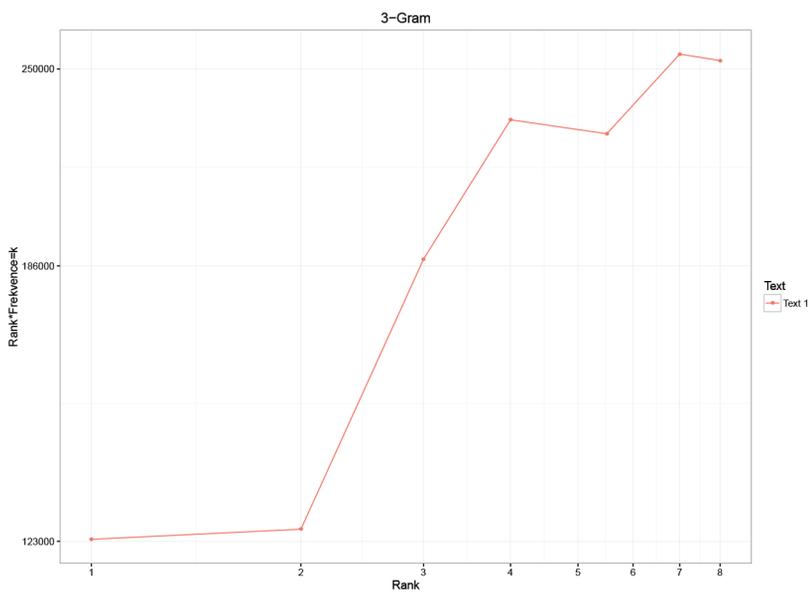
20-gram textů českého jazyka (text byl náhodně vybrán z korpusu cca 1000 českých textů)



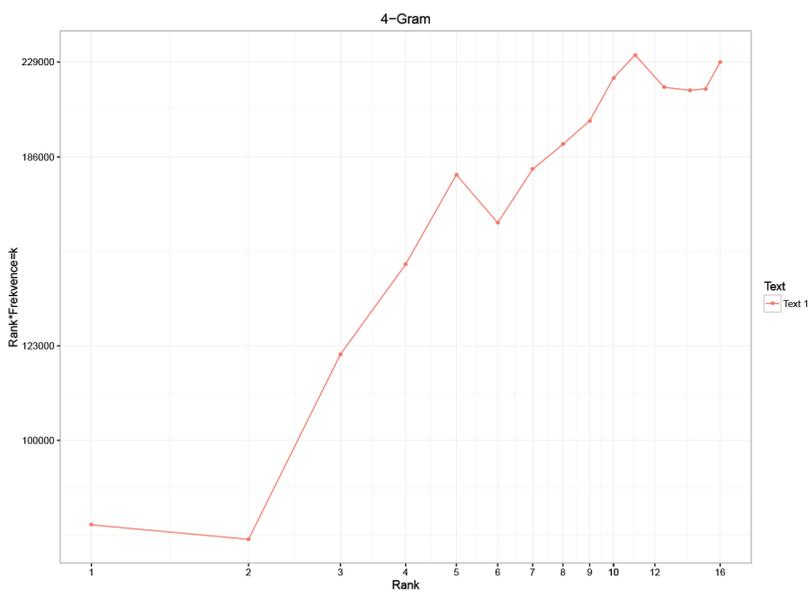
1-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



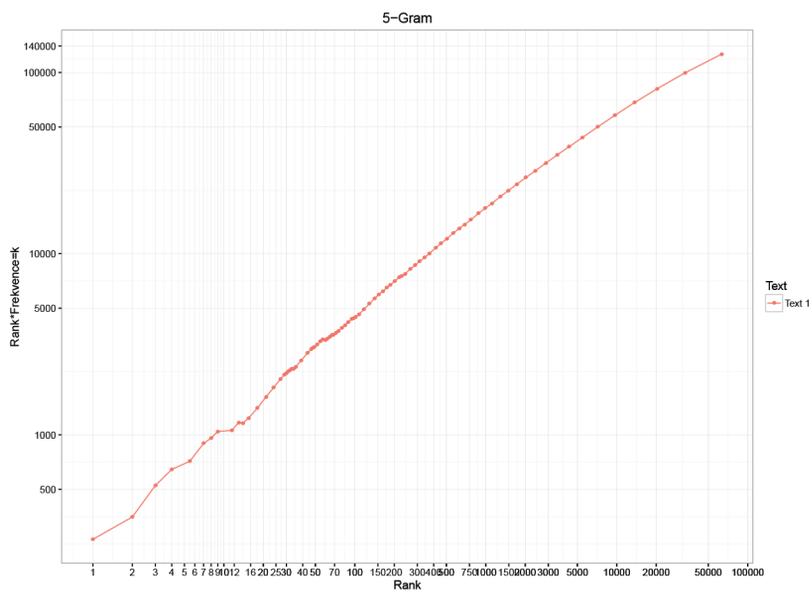
2-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



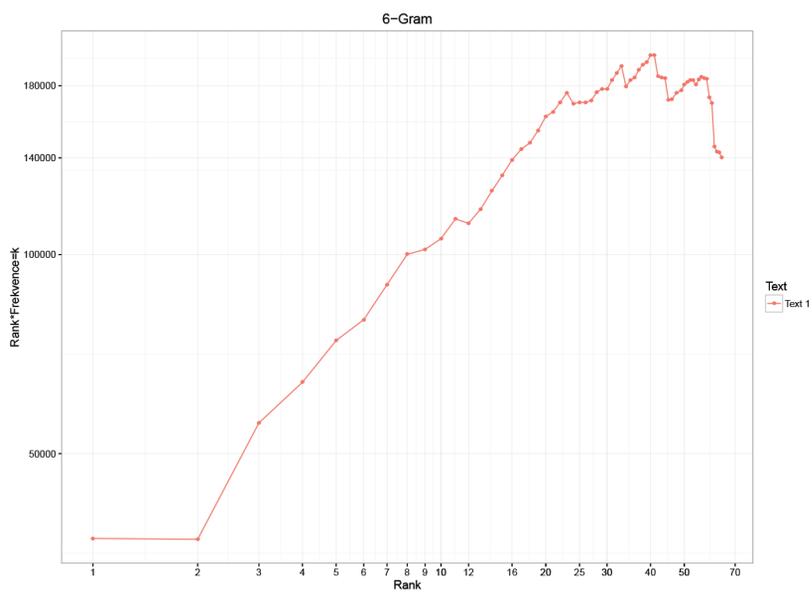
3-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



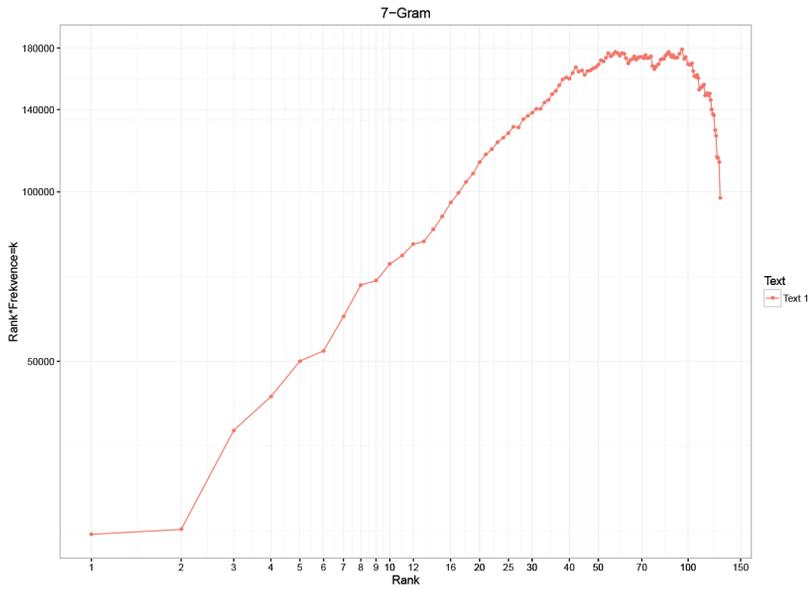
4-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



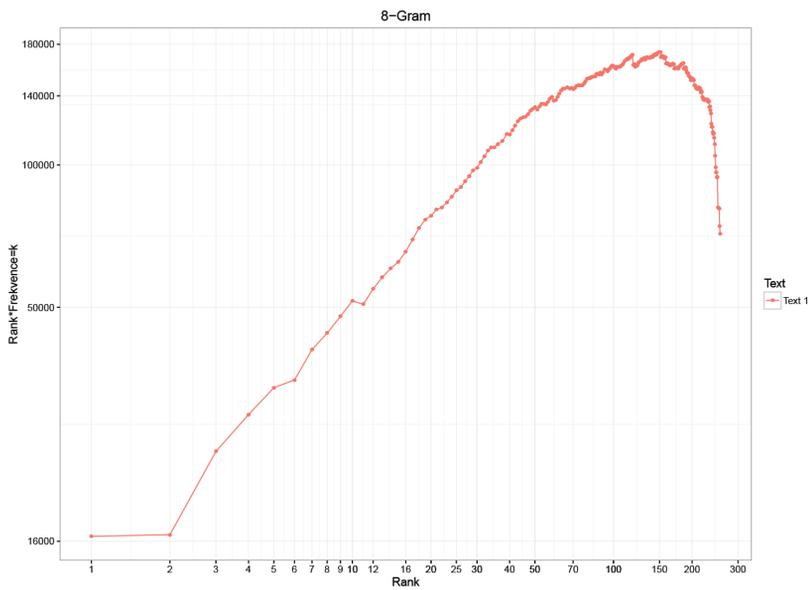
5-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



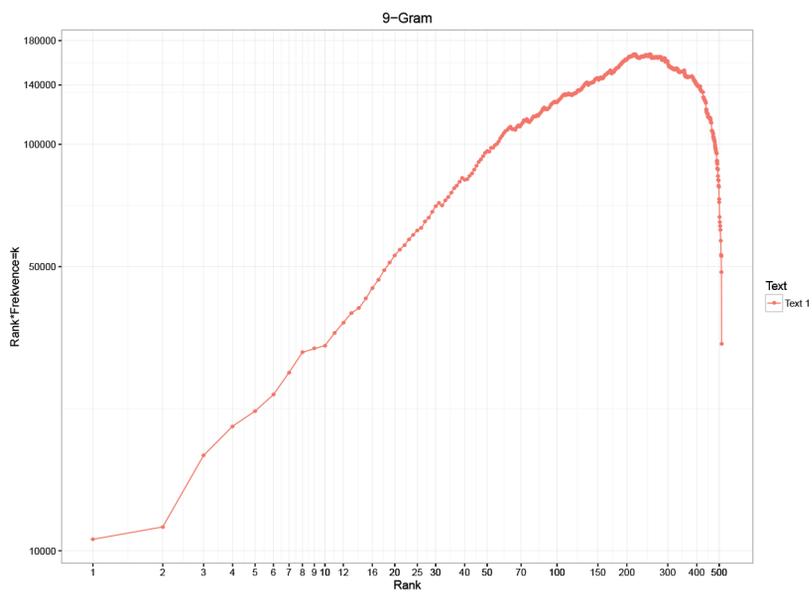
6-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



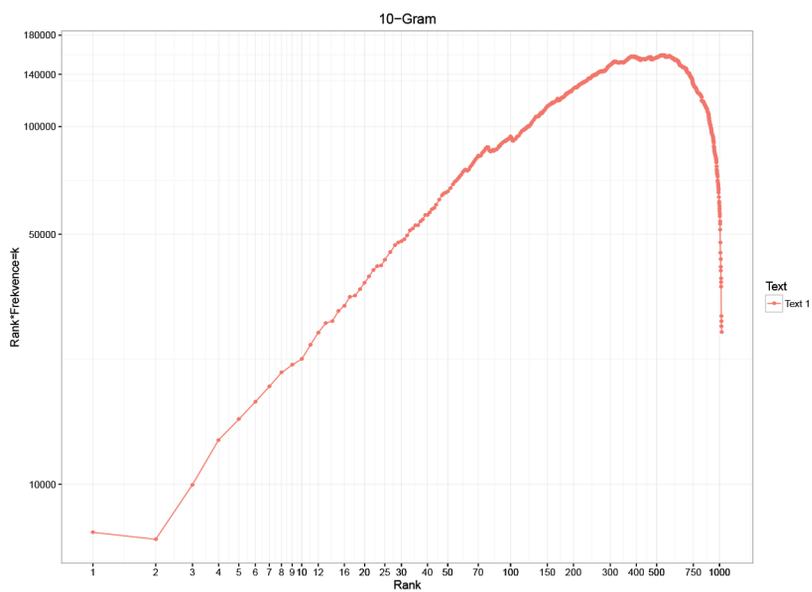
7-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



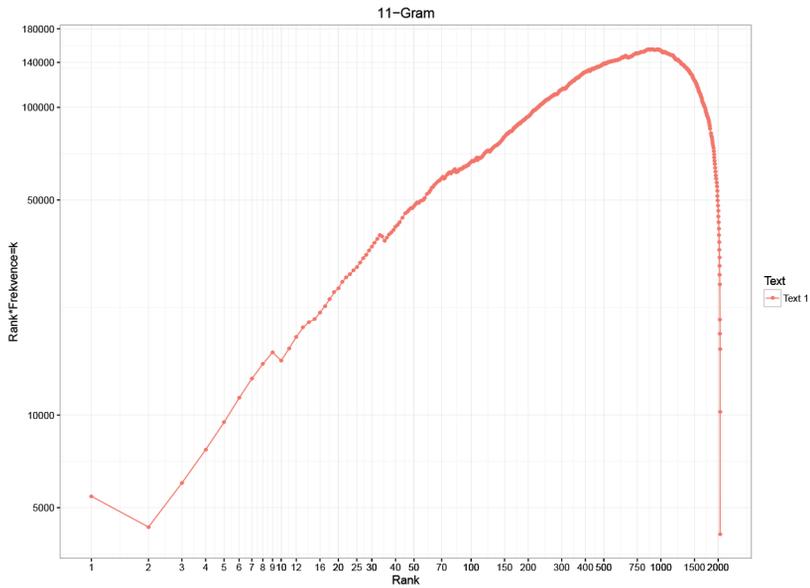
8-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



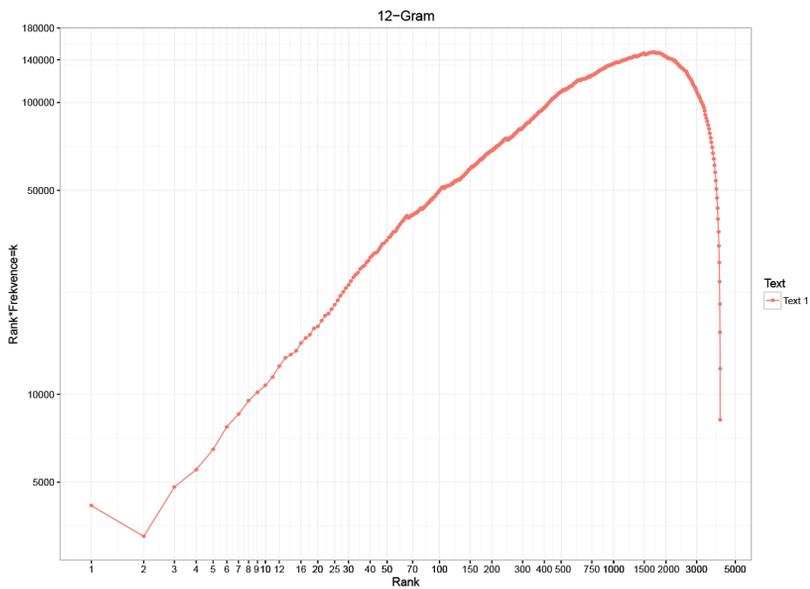
9-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



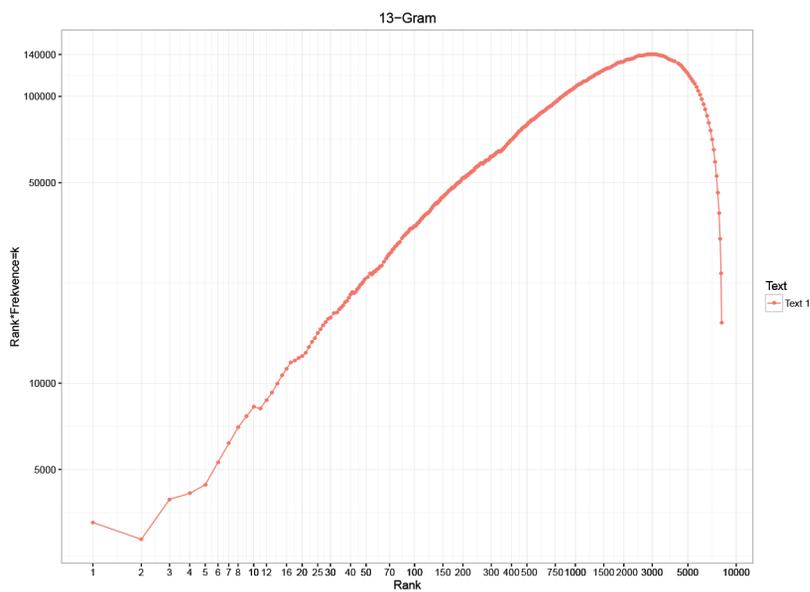
10-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



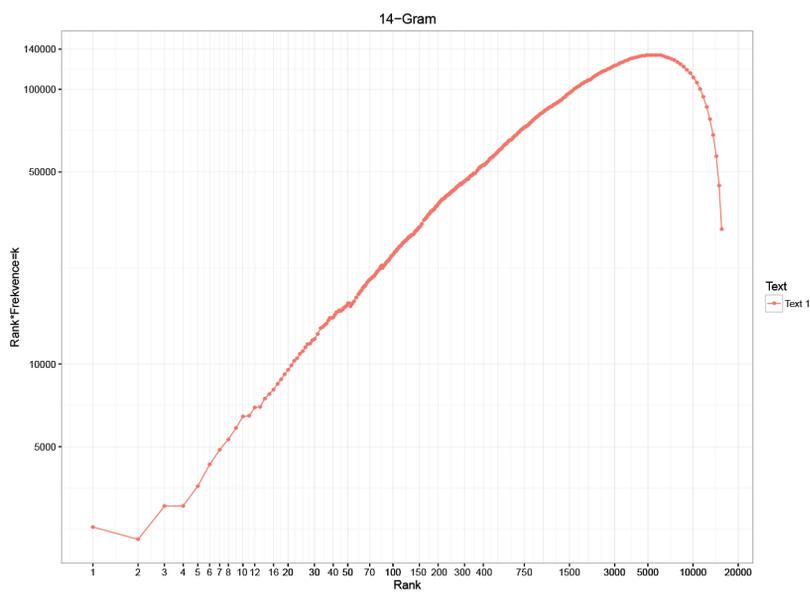
11-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



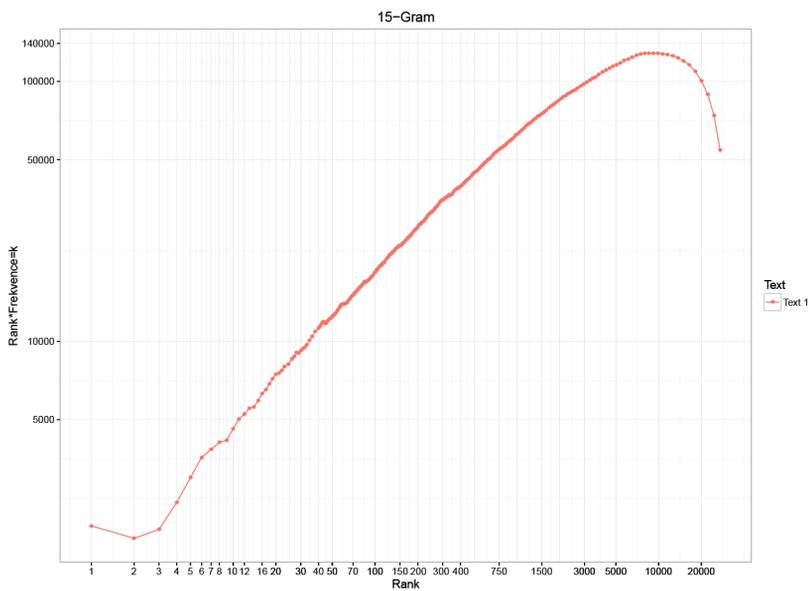
12-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



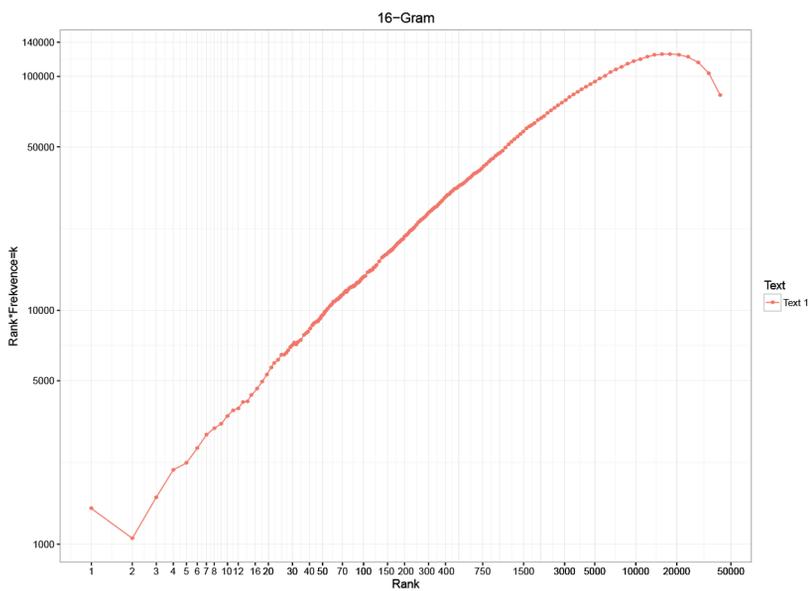
13-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



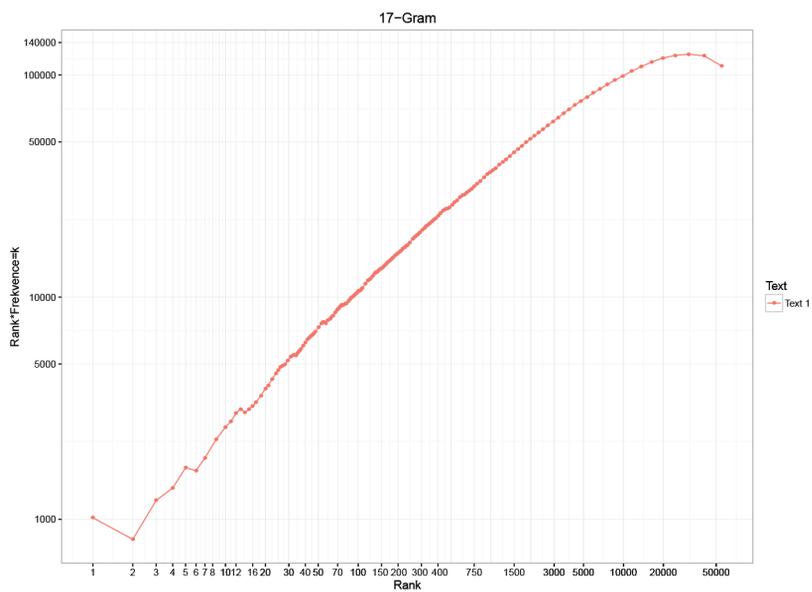
14-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



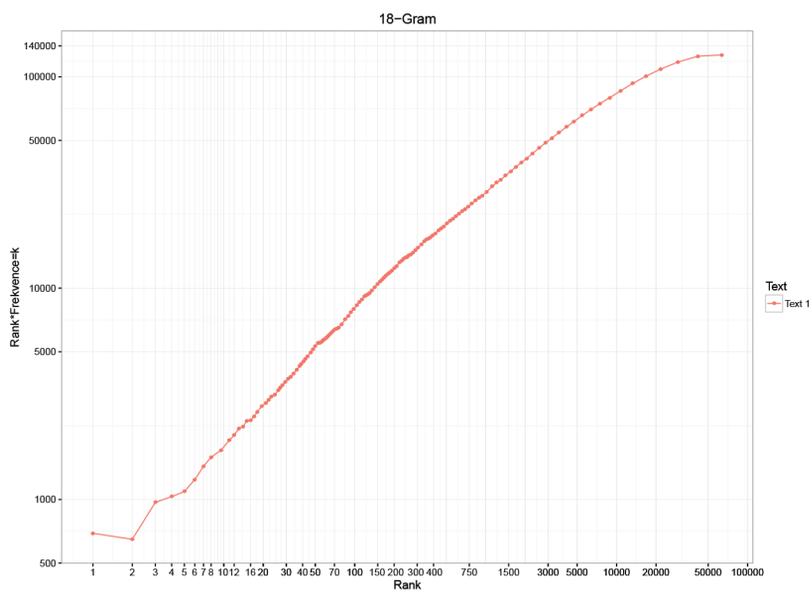
15-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



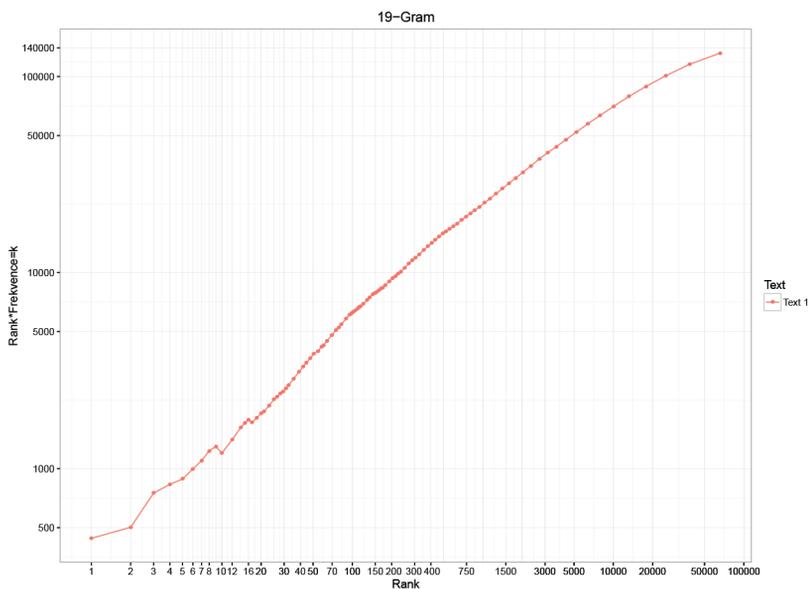
16-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



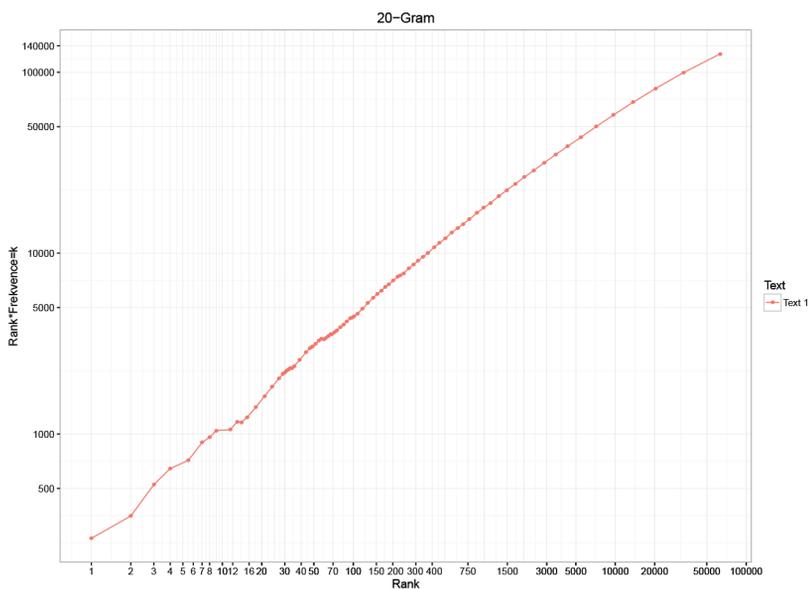
17-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



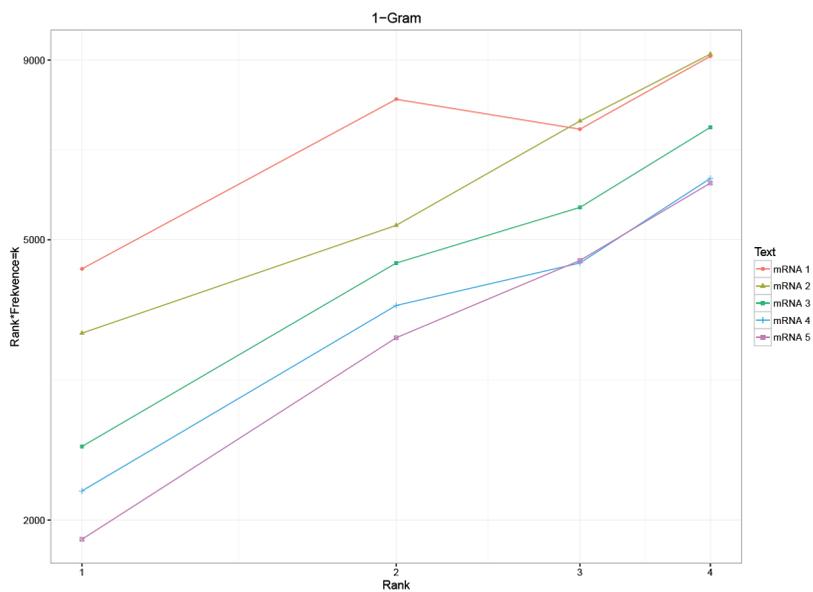
18-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



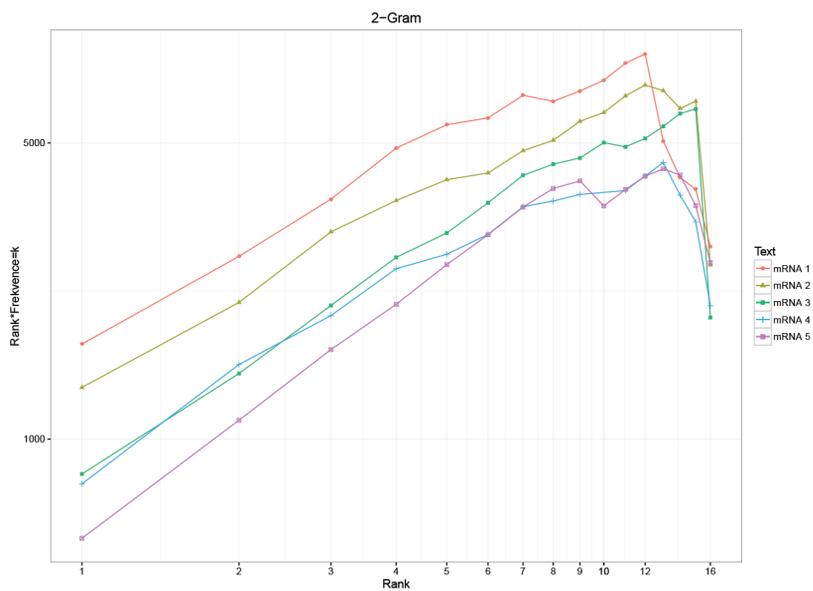
19-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



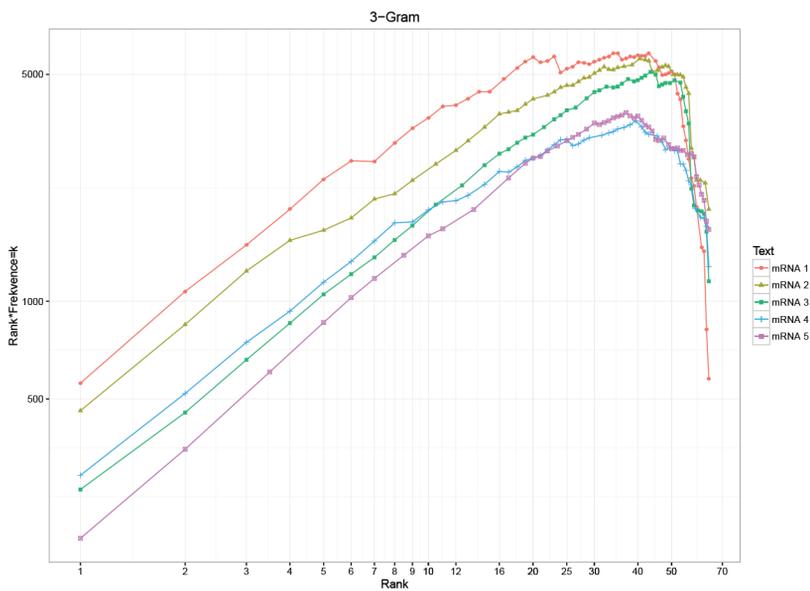
20-gram text čínského jazyka (text byl náhodně vybrán z korpusu cca 100 čínských textů)



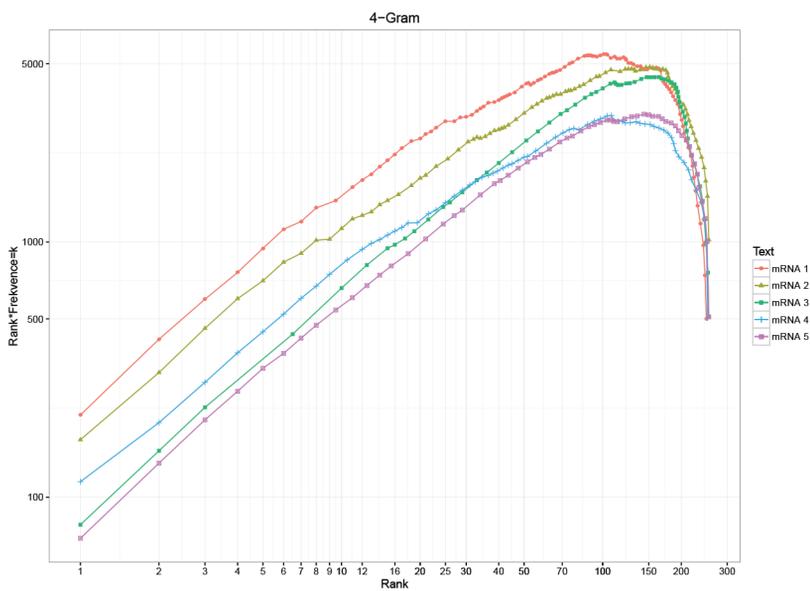
1-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



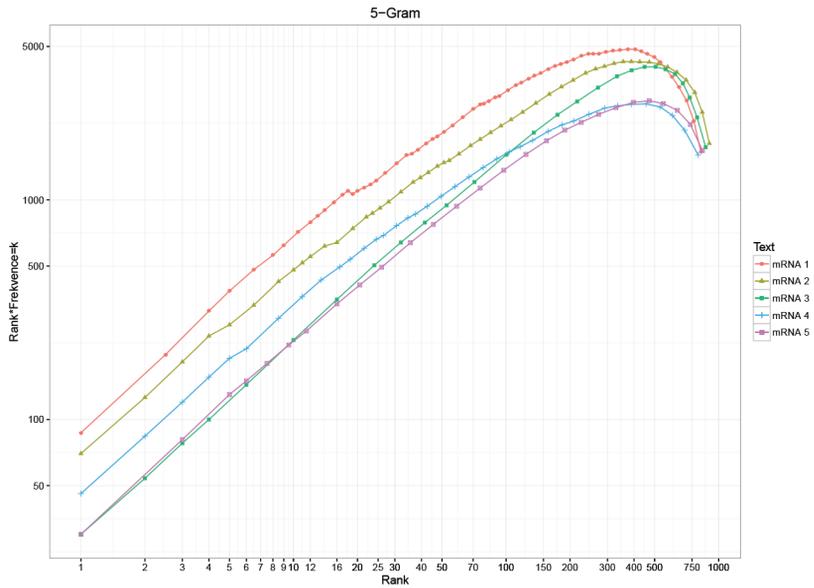
2-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



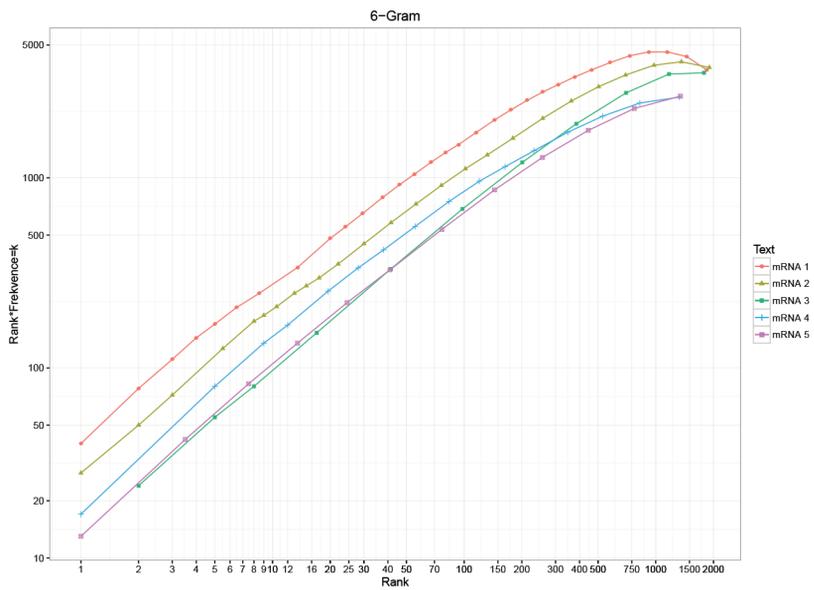
3-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



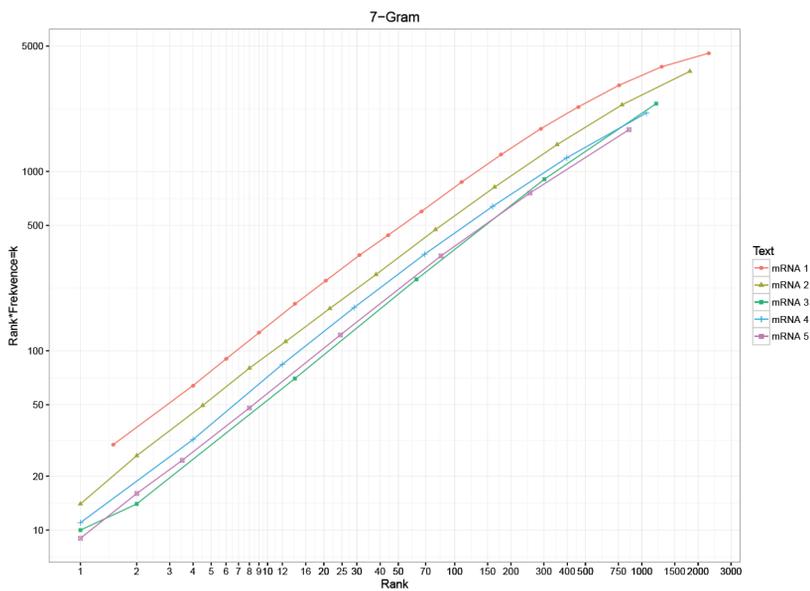
4-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



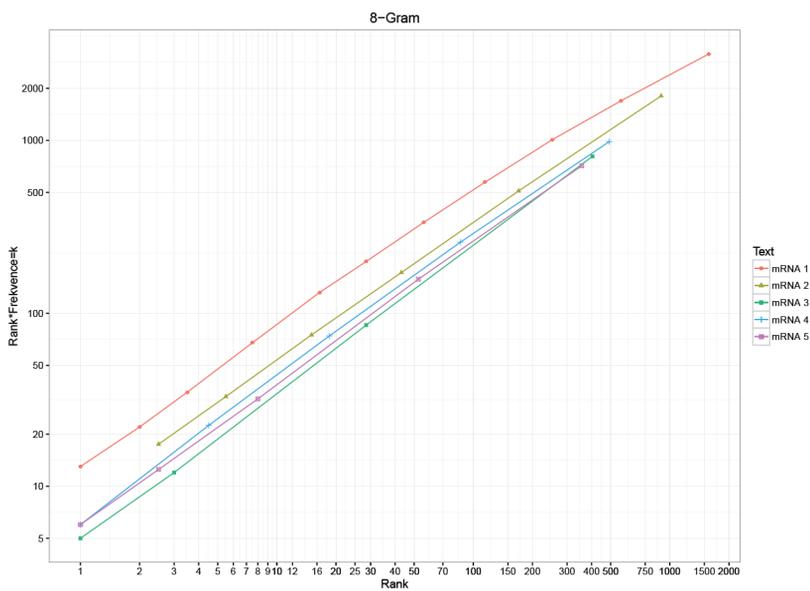
5-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



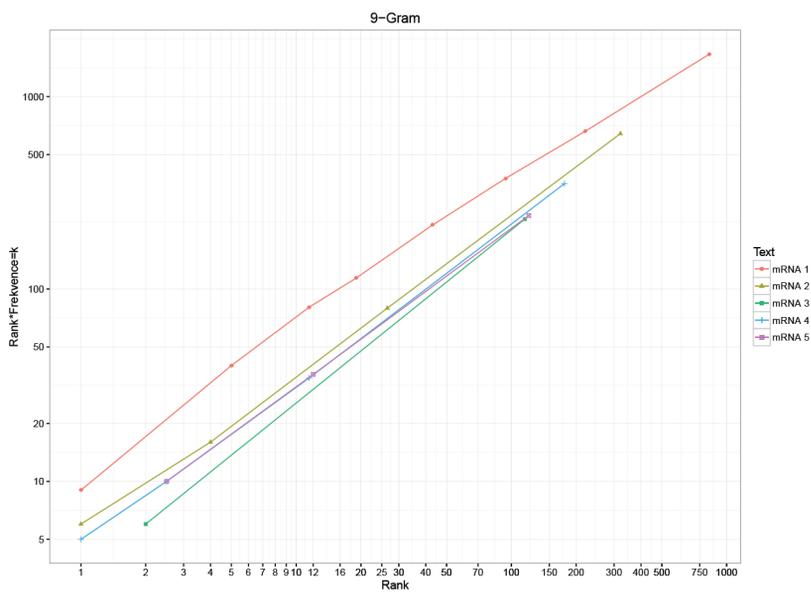
6-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



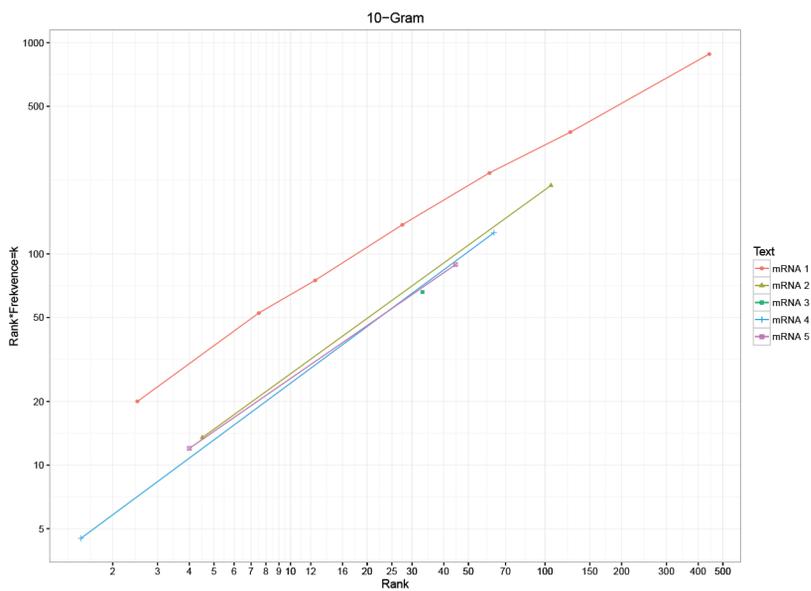
7-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



8-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



9-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)



10-gram mRNA (GenBank: AAA52700.1, AAA53189, AAA59925, AAA60577, AAA59204)

Věcný rejstřík

A

adaptor 26, 28, 31, 36

adjektivum 70

aminokyselina 23-28, 31-32, 36-38, 41, 52, 57, 59-60, 66-67, 73-74, 85

analogie 20-21, 25-28, 33, 38-42, 49, 52-54, 56, 58-62, 67, 85

arbitrárnost (znaku) 16, 21, 24, 26, 31, 52, 57-58, 66, 73, 85

artikulace 9-10, 12, 15-18, 20-21, 25, 28, 38

B

Balnibarbi 14-15, 21, 85

bioinformatika 74

biosémiotika 26, 28, 38, 66, 85-86

Č

částice 63

D

distinktivní rys 17-20, 27-28, 33, 51-54, 58, 85

DNA 21, 24, 26-29, 31-42, 49-61, 66, 69-73, 85, 87-88

dvojitá artikulace 16-17, 20, 38

E

ekonomizace/ekonomie 11, 14-18, 20, 22, 24, 58

F

foném 38, 64

fonologie 11, 22, 52, 56, 58

forma 34-35, 60

funkce 12, 15, 19, 21-28, 32-33, 36, 39-40, 42, 58-61, 65-68, 71-74, 85

G

genetický kód 20-21, 27, 31-33, 37-39, 41-42, 49, 52-54, 59, 71, 73, 85
genom 39-40
gramatika 38, 55-62, 64-71, 73, 85-86

H

hezitační zvuk 10
hláska 9-11, 14, 16-22, 24-25, 27-29, 57, 59-62, 65, 85
hnRNA 32, 66

Ch

chaperon 71

I

index 56
informace 19-20, 28-29, 32, 34, 38-40, 42, 59, 63, 74
interpretans 12

J

junk DNA 39-40

K

kodon 27, 29, 32-33, 37, 57
kódující DNA 39-41
komunikace 11-12, 14, 20, 27, 55-56, 60, 68, 70
komutace 22, 59, 64
konformace 23, 26, 32, 36, 67, 71

L

lexém 63-64, 69-70
linearita 57, 73
lingvistika 11, 33, 37-39, 42, 52, 56, 68, 85

M

mapování 46-49, 51-53
 metafora 15, 33, 38-40, 49, 56
 mluvnice 9, 56, 65
 molekulární biologie 24, 31-34, 38, 40, 66
 morfém 16, 19, 25, 64-65
 morfologie 19, 25, 70
 mRNA 24, 26, 32, 36-37, 49-52, 66, 85, 87

N

náhodný text 42
 nekódující DNA 39-41
 n-gram 41-42, 44-46, 53, 87

O

obsah 12-17, 19, 21-29, 31, 34, 40, 55-58, 60-73, 85-86

P

pád 64-65, 67, 85
 písmeno 10, 28, 31-34, 38, 43-49, 51-54, 58, 66, 73, 85
 popis gramatický 55-57, 61-62, 64-65, 67, 85-86
 predikace 69
 predikátor 62-64, 70
 protein 23-26, 28, 31-32, 35-36, 38-40, 50, 56-61, 66-67, 71-73
 předložka 65, 67
 purin 58-59
 purport 12-13, 22-23, 26
 pyrimidin 58-59

R

RNA 24, 35, 38, 59, 66, 87

Ř

řetězec 9, 18-20, 23-28, 32, 34-38, 40-42, 60, 66-68, 71, 73, 86

S

sémantická báze 55, 68-69, 71, 86

sestřih 71, 73

slabika 10-11, 18-19

sloveso 62-64

snRNA 32, 66

string 87-88

substantivum 62, 64-67

syntax 62, 70

Š

šum 17, 20, 57

T

text 9-10, 12, 16, 19, 21-22, 25-27, 31-35, 38-57, 59, 61, 66-70, 85-87

tón 10, 17-18, 57

transkripce 27, 35-36, 38, 60, 66, 71

translace 24, 36, 38, 60, 66

triplet 27-29, 31-33, 36-38, 41, 49, 52-54, 57-60, 66-67, 71, 74

tRNA 24, 26, 31, 36

tvár 12, 18, 23-25, 28, 32, 36, 40, 55, 65-67, 69-72

typologie 55-56

U

umwelt 12-13, 22-23

V

valence 63

věta 16, 22, 25-26, 33, 38, 41, 44-45, 49, 54-55, 58-60, 63-65, 67-70

výraz 12, 15-18, 21, 24, 26-29, 55-58, 60-61, 63-64, 66-74, 85-86

Z

Zipfův zákon 39-43, 85

znak 12, 14-16, 20-22, 26-29, 38, 42, 44, 48, 55-59, 63, 66-68, 71-73, 85-86

Jmenný rejstřík (hesla)

A

Alberts, B. 35, 37, 40

B

Bolshoy, A. 41

C

Collado-Vides, J. 38, 56

Crick, F. 31-32, 66, 85, 87

H

Hála, B. 10

Hartl, D., L. 33

Higgins, H. 9

Hjølmslev, L. 12-13

J

Jakobson, R. 33, 38, 49, 54, 58-60, 85, 87

K

Kořenský, J. 56, 68-69, 71

M

Mantegna, R., N. 39-41

Monod, J. 37, 57-58

P

Popov, O., S. 32

R

Raible, W. 33, 57

Ruvolo, M. 33

S

Searls, D., B. 33, 56

Segal, D., M. 32

Stanford, A., L. 32

Sweet, H. 9

T

Tolkien, J., R., R. 45

Trifonov, E., N. 25, 32, 38, 66

Tsonis, A., A. 42

Twyman, R., M. 33, 36

V

Viewegh, M. 43, 45, 47

von Uexküll, J. 12-13

W

Weaver, R., F. 32, 35-37

Z

Zipf, G., K. 43

KATALOGIZACE V KNIZE - NÁRODNÍ KNIHOVNA ČR

Faltýnek, Dan Gramatiky DNA: báze nejsou písmena / Dan Faltýnek, Vladimír Matlach. – 1. vydání. – Olomouc : Univerzita Palackého v Olomouci, 2014.
– 122 stran. – (Qfwfq ; sv. 18)

Anglické resumé

ISBN 978-80-244-4341-6

575.113.1 * 577.21 * 81'1 * 003.2/.6 * 57:003.2/.6 * 81'36 * (048.8:082)

- genetický kód
- DNA (nukleová kyselina)
- lingvistika
- sémiotika
- biosémiotika
- mluvnice
- kolektivní monografie

81 - Lingvistika. Jazyky (11)

Gramatiky DNA: Báže nejsou písmena
Dan Faltýnek, Vladimír Matlach

Edice Qfwfq, sv. 18

Výkonný redaktor: Jiří Špička
Odpovědná redaktorka VUP: Lucie Loutocká
Jazyková redakce a sazba: Zuzana Indrová
Obálka: Martina Šviráková
Sazba: EZ Centrum s. r. o.

Vydala a vytiskla Univerzita Palackého v Olomouci
Křížkovského 8, 771 47 Olomouc
www.vydavatelstvi.upol.cz
www.e-shop.upol.cz
vup@upol.cz

Olomouc 2014
1. vydání, 124 stran

ISBN 978-80-244-4341-6

Publikace je neprodejná
VUP 2014/0935