JIŘÍ LEVÝ

PŘEDBĚŽNÉ POZNÁMKY K INFORMAČNÍ ANALÝZE VERŠE

A. Veršový systém a kód

1.1 Teorie informace pokládá jazykový systém za kód a stylistický systém za speciální případ jazykového kódu (označovaný někdy jako subkód).[1] Nejde přitom jen o nahrazení pojmu „systém" termínem „kód", ale o jiný analytický přístup:[2] jazykové sdělení se zkoumá jakožto lineární sled diskrétních jednotek, které v nejvyhraněnějších případech mají binární charakter; toto pojetí umožňuje průběh řady přesně matematicky vyčíslit s použitím počtu pravděpodobnosti. Proto snad nebude bez užitku pokusit se z hlediska nové metody formulovat problematiku verše a zjišťovat, v čem matematické metody mohou zpřesnit metrickou analýzu.

„Jazykovým kódem rozumíme soubor jistých jazykových prvků a pravidla kombinování těchto jazykových prvků. Soubory základních prvků možno označit jeho inventáře kódu a soubory pravidel usměrňujících kombinování prvků jako schémata kódu."[3] Přitom platí, že a) každé *schéma*

[1] Užívání a užitečnost pojmu kód nejsou zcela vyjasněny a někdy se právě matematikové snaží tento pojem rozšířit i na oblasti, které zatím nejsou matematicky vystižitelné v kategoriích lineárního stochastického procesu: dávají mu pak význam jen matně odlišný od pojmů „metoda", „systém postupů" nebo pod.: „Jak jsme tedy svrchu poznamenali, naše teorie jazyka poskytuje vhodnou nomenklaturu pro výpravné umění. ‚Dramatický' a ‚scénický', ‚epický', ‚malebný' způsob předvádění, ‚autor sám' ‚autorovy výtvory' v rámci vyprávění a ‚postavy', to vše jsou různé kódy, z nichž každý předvádí život z jiného úhlu. Jejich potřebnost vzniká tím, že žádný z těchto kódů sám o sobě nemůže být práv složitosti společenského života lidí. Výhodou této aplikace jazykové teorie na výpravné umění není jen to, že poskytuje jasně definované pojmy za dosavadní popisné termíny užívané Lubbockem. Umožňuje nám porozumět některým uměleckým postupům..." (G. Herdan, *Type-token mathematics*, S-Gravenhage 1960, 274). Herdanovy analytické poznámky ke konkrétním literárním dílům však nepřinášejí poznatky, které by nebylo možno získat tradičními metodami.

[2] Srov. J. Levý, *Teorie informace a literární proces*, Česká literatura XI, 1963, 281—307.

[3] J. K. Lekomcev, *Zamečanije k voprosu o dvustoronnem jazykovom znake*, Voprosy jazykoznanija X, 1961, No. 2, 36—41. (V dalším budu odkazovat na tuto stať v textu zkratkou Lek.).

má svůj *inventář* prvků; b) inventář se skládá z určitého *počtu možných* prvků, z nichž každý se v jazykovém sdělení vyskytuje s určitou průměrnou *pravděpodobností.*

1.2 „Každé schéma je možno si představit jako pořádající princip vytvářející odpovídající jazykové jednotky. Základní operací tohoto pořádajícího principu by byl výběr určité kombinace prvků (mající u lineárních schémat podobu sledu prvků) z množiny možných kombinací... Schéma tedy zřejmě znamená rozestavení prvků v jednotce podle omezeného počtu pozičních tříd" (Lek. 39). To je přesný popis funkce rytmického schématu (a také rýmového schématu apod.): ze všech možných sledů prvků, t. j. slov, vybírá takové, které dávají odpovídající rytmické jednotky; např. v jambu po slově „překrásný" může následovat jen slovo jednoslabičné, po „krásný" jakékoliv slovo.

Schéma pořádá prvky kódu tím způsobem, že z možných kombinací vybírá některé, a tím současně v řadě prvků, které měly všechny stejnou pravděpodobnost, uděluje některým pravděpodobnost vysokou (těm, kterých užívá k sestavení jednotek), jiným nízkou (těm, které schéma vylučuje).

Jakým způsobem omezuje schéma inventář, sledoval pro ruský verš podrobně J. N. Golemyščev-Kutuzov[4] a dospěl k těmto závěrům:

V ruštině existují slova, v nichž po přízvuku může následovat, příp. před přízvukem předcházet, minimálně 7 slabik. Z celkového počtu 47 různých slovních typů 22 se v klasických rozměrech nemůže vyskytnout, 12 se může vyskytnout výjimečně, 13 tvoří základní slovní fond ruského verše. Mluví o „7 slabikách jakožto přirozené hranici slov, která se vyskytují v rytmickém systému ruštiny. Za hranicí 7 slabik začíná próza". Z hlediska počtu pravdepodobnosti je ovšem takové tvrzení příliš absolutní. Je třeba odlišovat dvojí omezení inventáře schématem:

a) prvky inventáře, které se příčí schématu, a proto jich absolutně nelze užít (to jsou výjimečné případy);

b) prvky, u nichž je vlivem schématu jen snížena frekvence — sem patří většina typů, jež Golemyščev-Kutuzov vylučuje (ostatně ani v próze se všech 47 typů neužívá běžně).

Schéma dále omezuje uspořádání prvků inventáře v řadě. Nepočítáme-li s omezením kombinačních možností vlivem metrických schémat, je v sedmislabičném ruském verši teoreticky možných 365 různých kombinací slov, v osmislabičném kolem tisíce, v desítislabičném několik tisíc. V osmislabičném jambu je však podle Golemyščeva-Kutuzova jen 41 reálných možností, v desítislabičném jambu nebo trocheji po 153 kombinačních možnostech (častěji se z nich užívá jen asi 100), v třístopém amfibrachu 9 a v pětistopém amfibrachu 81 variant. Z těchto údajů je zřejmé, že elimi-

[4] J. N. Golemyščev-Kutuzov, *Slovorazdel v russkom stichosloženii,* Voprosy jazykoznanija 1959, No. 4, 21 n. Omezení diferenciace rytmických typů slov ve verši má pochopitelně za následek snížení entropie slovního skladu (ačkoliv z některých jiných hledisek, především významového, má slovo ve verši vyšší entropii než v próze); propočítal je A. M. Kondratov, *Teorija informacii i poetika* (Problémy kibernetiky, vyp. 9, Moskva 1963, 279—286).

ační působení schématu je slabší u jambu a trocheje než u amfibrachu. Také omezení kombinačních možností schématem je však třeba chápat stochasticky, nikoliv deterministicky; kombinace

v v v v v v v ⏑ / v ⏑
v v v v v v v — v / ⏑,

které Golemyščev-Kutuzov označuje za nemožné, mají ve skutečnosti pouze mizivě malou pravděpodobnost výskytu.

1.3 „Schéma jednotky může vykonávat svou funkci jen tehdy, jestliže prvky, se kterými dané schéma operuje, budou mít nějaké příznaky, podle nichž může být prvek zařazen do určité poziční třídy schématu; např. ve schématu slabiky mohou jako takový poziční příznak sloužit některé momenty ve skladu foném z distinktivních příznaků, ve schématu věty mohou analogickou úlohu hrát některé gramatické morfémy ve slově. Je možno to formulovat tak, že každá jednotka existuje v kódu (na rozdíl od její existence v inventáři) jakožto souhrn informací o pozičně-kombinatorických vlastnostech dané jednotky a prvků různých rovin, které ji skládají. Každé schéma musí jednotce, kterou vytváří, udělit příznak potřebný pro uskutečnění dalšího schématu v hierarchii schémat" (Lek., 40). Toto pojetí upozorňující na potřebu lišit pozičně-kombinatorický příznak od fonetické (nebo jiné) kvality sdělení se shoduje s teoretickými distinkcemi vžitými v teorii verše, především s odlišením metrického důrazu proti fonetickému přízvuku apod. Schémata pořádající jazykové prvky do veršového útvaru jsou obyčejně složitá, jediné schéma pracuje se soustavou několika diferenčních příznaků. Např. na vytváření metrického schématu se účastní alespoň 2—3 příznaky: a) přízvuk (tj. např. jeho pravidelná alternace), b) slovní předěl (a jeho rozložení ve vztahu k přízvukům), příp. c) cézura.

Ústřední otázkou každé verzifikace je, které lingvistické prvky mohou sloužit jako distinktivní příznaky veršového systému. V této otázce se informačněteoretické stanovisko mnohdy stýká se stanoviskem fonologickým (funkčním): jazykový příznak, který je vázán na jiný příznak (např. český přízvuk na slovní předěl), je silně redundantní, má nízkou nebo nultou informační hodnotu, protože po objevení se primárního jazykového příznaku jej s jistotou očekáváme. Jazykový příznak, jehož distribuce není takto vázána, má informační hodnotu a zpravidla i fonologickou distinktivní platnost.

Z informačněteoretického hlediska je možno velmi jednoduchým způsobem vyvrátit námitku, kterou proti Jakobsonovu fonologickému výkladu rytmických příznaků uvádí de Groot: „Podle známého tvrzení Jakobsonova jsou v básni stylizovány jen distinktivní příznaky... Příkladem redundantního příznaku, který podle Jakobsona není nikdy stylizován, je pevný slovní přízvuk v jazycích, jako je čeština, maďarština a latina. Necháme-li stranou problémy verzifikace v těchto jazycích, můžeme říci, že

není zdá se žádný pádný důvod, proč by takové příznaky neměly nikdy být esteticky stylizovány, pokud jen jsou konvenční a konsistentní. Můžeme to ilustrovat srovnáním z jiného pole. V souboru 52 bridžových hracích karet jsou obrázky (piky, srdce, kříže, kára) distinktivní; barvy (černá a červená) nejsou distinktivní. Není žádný pádný důvod, proč bychom neměli hrát bridž s čistě černobílými kartami. Není však také žádný apriorní důvod, proč by pro estetické účely umělec neměl bridžové karty uspořádat současně podle kreseb a barev nebo dokonce jen podle barev."[5] De Grootova námitka je lichá a jeho analogie nepřiléhavá již z tohoto prostého důvodu: v množině karetní hry jsou barvy a figury aleatoricky rozloženy, nikde nejsou vázány figury jen na určitou barvu nebo barvy na určitou figuru, proto je možno obou příznaků užít současně k odlišení jednotek kódu; naproti tomu přízvuk a slovní předěl jsou na sebe v češtině vázány, tedy jeden z obou příznaků pozbývá informační hodnoty a není možno k rytmickému kódování užít obou současně. De Grootova argumentace současně ukazuje, jak je nebezpečná taková abstraktní kombinatorika.

1.4 Jazykové jednotky se teprve při kódování přetvářejí v jednotky kódu: „... jazykové jednotky, kromě diferenčních příznaků, fonémů a sémantických příznaků, nejsou prvky inventáře kódu, ale teprve dodatečně se v kódu konstruují" (Lek., 41). Také jazykové jednotky verše se dotvářejí teprve v procesu kódování; realizují se přitom zvláště ty příznaky, které jsou rozhodné pro poziční zařazení jednotky:

a) Po rytmické stránce se teprve při kódování rozhoduje o přízvučnosti či nepřízvučnosti monosylab a v delších slovech o přízvučnosti všech slabik počínajíc třetí; naopak součástí kódu nejsou některé reálné kvality jazykových jednotek (jako je např. odstupňování slovních a větných přízvuků).

b) V inventáři prvků rýmu se uplatňují jen ty diferenční příznaky, které jsou relevantní v rýmovém schématu, a nikoliv všechny ty, které jsou relevantní (fonologické) v jazyce: např. v českém rýmu některých období je irelevantní nejen protiklad znělosti souhlásek na konci slov, ale také protiklad kvantity u samohlásek, příp. protiklad souhlásek palatálních proti nepalatálním atd. — tedy mnohé fonologické příznaky se nestávají diferenčními příznaky schématu.

c) K velmi podstatné redukci dochází v instrumentaci; např. ve zvukové řadě zaměřené na labiální souhlásky, frikativy, přední nebo zadní samohlásky atd. stávají se ostatní příznaky irelevantními.

Jednotky daného jazykového kódu někdy fungují ve verši v redukované podobě, jindy v podobě obohacené o další diferenční příznaky vnucené jim schématem. Vzhledem k tomu, že literární kód, na rozdíl od lingvistického, je historicky silně proměnlivý, je do jisté míry proměnlivá i jeho soustava diferenčních příznaků: v současné době jsou např. spory, zda je pro umístění v rýmu relevantní souhláska na konci rýmové slabiky (tzv. useknuté rýmy).

[5] A. W. de Groot, *Phonetics in its Relation to Aesthetics (Manual of Phonetics*, Amsterdam 1957, 389).

Při dešifraci nastává opačný postup než při kódování: nevytvářejí se z prvků jednotky podle schématu, ale z jednotek se vytváří schéma, tj. z jazykových jednotek verše vnímá čtenář rytmus, rým atd., a to v různém stupni úplnosti a diferenciace podle své citlivosti a zkušenosti s veršem. Při vzniku verše bylo dáno schéma a jeho působením z jazykových prvků vznikaly jednotky sdělení. Při četbě jsou dány jednotky sdělení a schéma, které z nich při čtenářské konkretizaci vzniká, je labilní a do jisté míry individuální, protože se „abstrahuje" z jazykové řady, v níž schéma (např. jambické) není přesně splněno, a proto může často dojít k dešifraci podle několika schémat (daktylotrochej proti jambu); kromě toho vzhledem k různě silné realizaci diferenčního příznaku nemívají jazykové jednotky binární charakter (srov. např. různou intenzitu a stálost přízvuku slovního i větného na jednotlivých slabikách).

1.5 Zatím jsme se zabývali lineárními schématy, která vytvářejí jednotky kódu tím, že pořádají prvky do posloupných řad; sem patří většina prozodických schémat. I ve verši však hrají úlohu jednotky, jejichž sklad nemá lineární charakter; sem řadí informační teorie především fonémy a sémantémy a snaží se najít klíč k jejich rozkladu: „Pod vlivem výsledné redukce informačněteoretických problémů na binární rozhodování, pokouší se lingvistické bádání rozštěpit v daném jazyku také fonémy do simultánních svazků (bundles) dále nedělitelných distinktivních příznaků (distinctive features). Rozkládání fonémů, jež se označuje jako *analytická transkripce* nebo *analýza v komponenty,* se opírá převážně o ty vlastnosti signálů, které je možno zjistit analýzou v rovině času a frekvence, a teprve v druhé řadě o artikulační vlastnosti".[6]

Jakobsonova a Halleova[7] rozkladu fonémů na 12 distinktivních příznaků se pokusili někteří autoři užít k měření vzdálenosti mezi jednotlivými fonémy;[8] do souvislosti s těmito vzdálenostmi byla uváděna také otázka přesnosti rýmu.[9] Zdá se však, že pro měření rýmových souzvuků nejsou distinktivní příznaky vhodné, protože a) nevystihují akustickou podobnost zvuků pro vnímatele; zjistilo se např., že v angličtině se *p* při poslechu častěji zaměňuje s *k,* od něhož je vzdáleno 4 d. p., méně často s *t* (3 d. p.), ještě řidčeji s *b* (2 d. p.)[10]; b) v rýmu hraje zřejmě závažnou roli právě artikulační hledisko — jako slabší odchylka od zvukové shody se např.

---

[6] W. Meyer-Eppler, *Grundlagen und Anwendungen der Informationstheorie,* Berlin-Göttingen-Heidelberg 1961, 319.

[7] R. Jakobson, C. G. M. Fant, M. Halle, *Preliminaries to Speech Analysis,* Cambridge 1952.

[8] Sol Saporta, *Frequency of Consonant Clusters,* Language 31, 1955, 25—30.

[9] Meyer-Eppler, cit. kn., 326.

[10] Tamtéž, 360.

pociťuje protiklad homorgánní souhlásky znělé a neznělé než např. *f* a *v*, *v* a *b*, *k* a *č*, mezi nimiž je také (aspoň v angličtině) rozdíl jen 2 d. p.

Dějí se také pokusy o analýzu sémantémů na diferenční příznaky. Zjišťují se sémantické prvky v intonaci[11], mimice,[12] apod. a kromě toho se pracuje na segmentaci a matematickém formulování oblasti sémantična obecně. Tyto práce jsou však zatím ve stadiu hypotéz a také se nedotýkají specifických otázek prozódie.

1.6 Výhodu informačněteoretického přístupu k otázkám verše bych viděl především v tom, že otevírá cestu matematickému precizování některých rysů veršové stavby, zjišťovaných zatím spíše subjektivním odhadem, a hlavně v tom, že vyžaduje užívání moderní stochastické matematiky, která je adekvátnějším nástrojem pro popis průběhových změn a procesů než stará deterministická matematika aplikovatelná spíše na statické a izolované jednotky. Pojetí verše jako řady jazykových signálů řízené dialektikou nutného a náhodného a analyzovatelné podle pouček teorie pravděpodobnosti má proti tradičnímu statistickému přístupu tu výhodu, že umožňuje zkoumat také vztahy mezi dvěma nebo více body ve struktuře, a tím učinit první krok k analýze struktury jako dynamického, průběhového systému.[13]

Na teorii verše se pokusil noetickou axiomatiku počtu pravděpodobnosti aplikovat G. Herdan: „... rým a rytmus ... spoutávají myšlenku, a podřídit vyjádření myšlenky tlaku takových formálních činitelů, dokonce možná poněkud deformujícímu, zdánlivě jedině proto, aby po určitém počtu slabik bylo znovu slyšet tentýž zvuk, vypadá na první pohled jako zrada na pravidlech logiky ... Ale navzdory této prozaické kritice verše poezie za všech dob a u všech národů vykonávala a vykonává velký vliv na lidskou mysl právě těmito myšlenku deformujícími rysy, které jsou čistému rozumu cizí. Co je toho příčinou? Začněme poznatkem, že dobré verše svým zdůrazňujícím účinkem vyvolávají dojem, že myšlenka byla předurčena již jazykem samým a básník ji potřeboval jen vyhledat (Schoppenhauer). To ale znamená, že rým a metrum působí jako náhražky za logický důkaz, nebo při nejmenším za racionální úvahu. Jsou projevem ‚samokontroly' (Eigencontrol), již jazyk jakožto forma vykonává nad ‚vyjádřením', a již je možno přirovnat ke kontrole, kterou vykonávají v próze rozum a logika nad ‚obsahem'."[14]

---

[11] Viz např. práce sovětského psychologa N. I. Žinkina.

[12] Viz pokusy s tzv. kinesikou (o tom stručná informace v časopise Divadlo 1961, 719).

[13] Moderními statistickými metodami začali pracovat na analýze rytmu někteří slovenští badatelé, zvláště F. Miko, G. Altman, R. Štukovský (Litteraria VI, Bratislava 1963).

[14] G. Herdan, *Type-token mathematics*, S-Gravenhage 1960, 278.

Soustavný výzkum otázek verzifikace metodou informační teorie nebyl dosud zahájen;[15] kromě zjištění, která prozodická fakta odpovídají základním kategoriím teorie informací, by v první etapě takového výzkumu bylo zapotřebí uvažovat o tom, jak je k analýze jednotlivých složek možno používat exaktních metod již vžitých a ve kterých ohledech by je bylo možno zpřesnit použitím pouček stochastické matematiky.

B. Prvky a schémata

Foném — zvukosled

2.1 Nejmenší lineárně uspořádatelné prvky verše, stejně jako jazykového sdělení vůbec, jsou fonémy. Podle toho, kolik rozdílných fonémů má jazyk ve svém inventáři, může a) vytvářet větší či menší počet hláskových kombinací určitého rozsahu, t. j. „slov" (to má dosah pro rým), b) více či méně „pestře" zvukově diferencovat jazykovou řadu (objektivní mírou pro tyto hodnoty je entropie).

Hláskový sklad jazykového sdělení je nejpestřejší tehdy, jsou-li v něm fonémy rozloženy náhodně, stochasticky, t. j. tehdy, když se v dostatečně dlouhém textu vyskytují všechny stejně často. Míra neuspořádanosti takové řady o n různých fonémech (maximální entropie) je pak dána vzorcem $H_0 = ld\ n$. Ve skutečnosti ovšem nejsou všechny hlásky stejně frekventní, míra neuspořádanosti hláskového repertoáru v jazykové řadě (entropie 1. řádu, H_1) je menší. Uvedu hodnoty H_0 a H_1[16]:

	H_0	H_1
čeština:	5,39	4,67
ruština:	5,00	4,35
francouzština:	4,92	4,16
němčina:	4,76	4,10
angličtina:	4,76	4,03
španělština:	4,76	3,98

Je tedy zřejmé, že z hlediska „pestrosti" hláskové řady je čeština v této řadě jazyků na 1. místě, španělština na posledním.

Průměrný hláskový sklad a „normální" seřazení hlásek bývají ve verši různým způsobem pozměněny. I. Fónagy[17] predikcí maďarských básní

[15] Některé pojmy nové metodologie aplikoval na verš M. Janakijev, *B'lgarsko stichoznanije*, Sofia 1960.

[16] Viz. L. Doležel, *Předběžný odhad entropie a redundance psané češtiny*, Slovo a slovesnost XXIV, 1963, 173 (hodnoty entropie jsou zde bohužel počítány z písmen).

[17] I. Fónagy, *Informationsgehalt von Wort und Laut in der Dichtung*, Poetics, Warszawa 1961, 590—605.

zjistil, že mezní entropie hláskového skladu ve verši je vyšší než v próze; k témuž závěru dospěl v cit. studii L. Doležel pro Kunderovy *Monology:* průměrná hodnota mezní entropie u jiných stylů se pohybovala mezi hodnotami 0,7034—1,7126, u veršů činila 1,9115. Přitom entropie 1. řádu byla právě u veršů nejnižší: 4,5722 proti 5,5919—4,7044. Rozdíl $H_1—H_\infty$ činí u poezie 2,6607 proti 2,9654—3,8885 u jiných stylů. Interpretujeme-li tyto hodnoty, můžeme říci o tendencích hláskového uspořádání:

a) Nízká entropie 1. řádu svědčí o tom, že řídké hlásky (písmena) jsou v poezii ještě řidší než v próze a že převaha nejčastějších hlásek je vo verši ještě výraznější. Po této stránce je tedy hláskový obraz verše méně pestrý; jinak je tento jev možno formulovat také tak, že verš staví spíše na hláskách typických a potlačuje hlásky netypické.

b) Vysoká mezní entropie svědčí o tendenci k tomu, aby hlásky následovaly za sebou v pravidelnějších intervalech než v próze, aby tedy byly poměrně rovnoměrně v řadě signálů rozloženy; to je statistický důsledek stylizace „plynulé", „harmonické" apod Od tohoto „normálního" pozadí, které jsme tradičními analytickými metodami neměli možnost zjišťovat, se ještě ostřeji odrážejí hlásková opakování (zvukosled, rým); tato výjimečná opakování jsou sice při četbě nápadná, ale nemají podstatný vliv na celkový statistický obraz hláskového skladu.

2.2 Moderními statistickými metodami je možno zkoumat nejen typy a frekvenci hláskových opakování, ale také charakterizovat, jak je naopak hláskové pozadí těchto zvukosledů diferencováno, „modulováno", abychom užili termínu P. Guirauda: „Pod pojmem modulace bychom chtěli studovat vztahy mezi průvodními hláskami. Tato hlásky mohou být shodné nebo více či méně odlišné. Postupným opakováním zvuků stejného timbru vznikne verš slabě modulovaný; a naopak. To je jiný princip než zvukosled (tonalité). Slova étincellement a flamboient mají jedno jasnou a druhé temnou instrumentaci; a obě dvě jsou málo modulována. Naopak éblouissement je silně modulováno rozestupem mezi třemi následujícími samohláskami *é, ou, i.* ... je možno připustit, že umírněná a čistá modulace vytváří jeden z půvabů jazykového projevu a jeden z pramenů libozvučnosti."[18]

Guiraud např. propočítal, které hlásky se opakují ve Verlainově verši řidčeji, než by odpovídalo statistickému průměru: „Naše tabulky ukazují,

1. Že verš je modulován, tj. že počet kombinací stejných souhlásek za sebou je daleko nižší (asi poloviční), než by se podle pravděpodobnosti dalo předpokládat; a současně že se tato modulace liší podle kategorie a místa artikulace souhlásek.

2. Tato modulace je velmi silná mezi likvidami, slabá mezi nazálami a střední mezi okluzívami."[19]

---

[18] P. Guiraud, *Langage et versification,* Paris 1953, 96.

[19] Tamtéž, 99.

Odhadem je možno zjistit jen velmi nápadné a mechanické zvukové figury. Propočtem pravděpodobností přechodu se dají zjistit i jemné korelační nebo disimilační tendence. Např. B. F. Skinner zjistil tyto tendence v samohláskové instrumentaci u Swinburna: „Jediná kombinace, u níž skutečná frekvence převyšovala očekávanou frekvenci o více než trojnásobek průměrné chyby bylo *ei* následované samohláskou *e*... byla zřejmá tendence, aby po *i* nenásledovalo *ei* a po *u* nenásledovalo *i*".[20]

Takové tendence nejsou bez estetických důsledků, lze jimi vysvětlit část „neopakovatelné", „osobní" melodičnosti verše jednotlivých autorů, ony jemnosti, které tradiční literární věda vydávala za „neanalyzovatelné" a nedefinovatelné.

2.3 Proti tradičním metodám, které byly schopny zjišťovat jen nápadné nakupení stejných hlásek na jednom místě textu, příp. v pozicích téže třídy, je možno s použitím pravděpodobností přechodu analyzovat i zdánlivě neorganizované uspořádání hlásek. Jsou možné dva krajní typy „normálního", „průměrného" rozložení jazykových (a nejen jazykových) prvků v literárním díle: A. víceméně rovnoměrné rozptýlení prvků v intervalech pohybujících se kolem střední hodnoty; B. nerovnoměrné rozptýlení, jehož střední interval může být stejný jako v případě A, ale jehož jednotlivé skutečné intervaly kolísají ve velmi širokém rozpětí (teoreticky 0—∞) podle frekvence příslušející jim ve stochasticky uspořádané řadě. Uspořádání podle A je idealizované, tendence k takovému uspořádání (která se ostatně ve verši zpravidla aspoň ve slabé míře projevuje) bude příčinou toho, proč mnohé verše plynou „hladce", „melodicky", „harmonicky" atd., ačkoliv v nich není na první pohled patrna žádná organizace hláskové řady. Ostatně toto odlišení má širší noetickou platnost: idealizace proti naturalistické skutečnosti, aristotelské pojetí pravděpodobnosti proti pojetí nearistotelskému apod.

Náhodné uspořádání nemusí být vždy esteticky *bezpříznakové,* i seskupení hlásek, které vzniklo zcela náhodně, může esteticky působit. Např. pravděpodobnost, že dva za sebou následující půlverše budou začínat na souhlásku *t* (za předpokladu, že *t* je zhruba stejně frekventní na počátku jako uvnitř slova) je

$$P_{t+t} = P_t . P_t = P_t^2 = 0{,}0522^2 = 0{,}0026.$$

To znamená, že takové spojení můžeme očekávat asi jednou v 400 verších, a odpovídá stochastické zákonitosti, že v Máchově *Máji,* jehož rozsah je přes 800 veršů, se objevil verš

Tiché jsou vlny, temný vod klín.

[28] B. F. Skinner, *A Quantitative Analysis of Certain Types of Sound-Patterning in Poetry,* The American Journal of Psychology LIV — 1941, 76.

Přesto působí toto spojení neočekávaně, aktualizovaně. Čtenář totiž nevnímá současně celých 824 veršů *Máje* jako statistický celek, nýbrž jen velmi malý výsek textu, v jehož rozmezí jsou taková řídká spojení neočekávaná. Proto se v těchto případech nekryje matematická „neočekávanost" s neočekávaností estetickou; zákon velkých čísel má při vnímání textu omezenou platnost. Vedle toho působí také významový paralelismus.

2.4 Literární teorie ví již řadu desítiletí, že estetická působivost některých složek literárního díla je závislá na tom, jak je daný prvek „normálně" častý, tedy s jakou pravděpodobností můžeme očekávat jeho výskyt, jaká je jeho *prediktabilita:* „... o eufonickém uplatnění jisté hlásky nerozhoduje jen počet opakování, nýbrž i poměrná její hojnost ve srovnání s frekvencí normální."[21]

Informační teorie dále zjemňuje tento poznatek tím, že upozorňuje, že míra neočekávanosti hlásky nezávisí jen na její absolutní průměrné frekvenci, ale také na pravděpodobnosti výskytu navozené předchozími hláskami, tj. na *pravděpodobnosti přechodu* mezi danou hláskou a hláskami ji v textu předcházejícími: „Např. v ruštině za zvukem č nemůže v témže slově následovat *s, c, f.* Nejpravděpodobnější ze souhlásek je *t.* Po spojení *čt* je nejpravděpodobnější samohláska *o* (čto), řidčeji se vyskytuje *e* (čtenije), ještě řidčeji *i* (čtica, čtit), zcela málo použitelné je spojení *čty* (mečty)."[22] To ovšem současně znamená, že překvapivost další hlásky je podmíněna nejen její průměrnou frekvencí, ale také hláskami, které předcházejí.

Pravděpodobnosti přechodu a od nich závislá entropie dodávají tím, že jsou ve verši (a jazykové řadě vůbec) rozloženy nestejnoměrně, některým pozicím a prvkům větší stavební závažnost než jiným: „Podrobnější pozorování ukazuje, že v německých textech — a to obdobně platí i pro jiné jazyky — připadá na *souhlásková* písmena vyšší informační hodnota než na *samohlásková* písmena a počátkům slov vyšší informační hodnota než *koncům* slov. Naproti tomu má konec věty zpravidla vyšší informační hodnotu než začátek věty."[23] Tyto poznatky ukazují, že různé typy zvukových opakování nejsou rovnocenné. Souhlásková instrumentace se opírá o hlásky s větším informačním zatížením než instrumentace samohlásková; s tím souvisí snad i fakt, že významově funkční instrumentace (onomatopoje) se opírá především o souhlásky. Náslovná aliterace staví právě na jazykových prvcích o největším *informačním* obsahu: a) na souhláskách, b) na začátcích slov. Naproti tomu rým se opírá o slaběji zatížené konce slov a v češ-

[21] J. Mukařovský, *Kapitoly z české poetiky II,* Praha 1948, 76.
[22] E. I. Šendels, *Svjaz jazykoznanija s drugimi naukami,* Moskva 1962, 55.
[23] W. Meyer-Eppler, *Grundlagen und Anwendungen der Informationstheorie,* Berlin-Göttingen-Heidelberg 1961, 354.

tině nadto o informačně méně funkční samohlásky; ovšem jeho *sémantická* hodnota je proti aliteraci zvýšena tím, že se nekryje jen jeden foném, ale zpravidla nejméně celá slabika, tj. jazykový segment potenciálně obdařený významem (slovní kořen, suffix atd.).

Slabika otevřená — zavřená

3.1 Jemnějšími statistickými metodami, které přinesla moderní teorie pravděpodobnosti i do literární vědy, se možná objeví zákonitosti uspořádání některých dalších jazykových složek verše. Polská badatelka M. R. Mayenowa má zásluhu, že začala zkoumat rozložení otevřených a zavřených slabik ve verši. Zjistila, že v *Panu Tadeuszovi* průměrná frekvence otevřených slabik je vyšší před cézurou (65,4 %) a na konci verše (70,7 %) než na konci slov uvnitř verše (60,7 %).[24] Vyvozuje z toho, že charakteristické tendence polské slabiky k otevřenosti je využito v uzlových bodech verše. V další etapě by bylo poučné srovnat procento otevřených slabik na konci syntagmat spadajících dovnitř verše a na konci syntagmat spadajících do uzlových bodů verše, aby se zjistilo, pokud je zvýšení dáno lokalizací syntaktických konců v uzlových bodech verše a pokud ještě jinými činiteli. Zjišťoval jsem v blankversech Hálkova *Záviše z Falkenštejna* (a to vždy pro 100 případů z 1.—4. dějství) rozložení otevřených a zavřených slabik a) na konci všech slov, b) na konci syntaktických celků (tj. před jakýmkoliv interpunkčním znaménkem nebo před souřadící spojkou spojující dvě věty), c) na konci veršů:

	Konce slov		Konce vět		Konce veršů	
	ot.	zav.	ot.	zav.	ot.	zav.
I	60	40	52	48	46	54
II	48	52	45	55	46	54
III	55	45	44	56	39	61
IV	58	42	40	60	41	59
Průměr	55	45	45	55	43	57

Na konci Hálkova blankversu je frekvence otevřených slabik (43 %) nižší než uvnitř verše (55 %) a to zřejmě v souvislosti s tím, že výskyt otevřených slabik je obecně poněkud snížen na konci syntaktických celků. Poměry v češtině, která má také tendenci k otevřeným slabikám, budou patrně poněkud jiné než v polštině; frekvence otevřených slabik bude

[24] M. R. Mayenowa, *Wiersz* (Rytmika, Warszawa 1963), 9.

patrně výrazněji kolísat na konci syntaktických celků než na koncích veršů.

3.2 Pokud se prokáže, že rozložení slabik podle otevřenosti není v českém verši náhodné, ale že je stylizováno podle místa ve verši, autora atd., pak by to znamenalo, že jde o jev stylistický, a bylo by třeba zjišťovat také jeho estetický význam.

Při takové distribuční analýze nejde jen o to zjistit, ve kterých bodech je jeden z obou protikladných typů nahromaděn, ale také vyšetřit, zde nejsou po této stránce mezi jednotlivými uzlovými body verše korelační vztahy. Jako příklad takového šetření uvedu statistiku otevřených slabik před cézurou (tj. ve 4. slabice) a na konci verše (8. slabika) Holečkovy *Kalevaly;* bylo excerpováno vždy po 100 verších z 10 zpěvů:

	4. slabika	8. slabika
I	71	72
II	66	69
III	68	73
IV	81	65
V	86	76
VI	83	71
VII	75	65
VIII	65	73
IX	88	74
X	75	76
Průměr	75	71

Bylo by možno klást otázku, zda je nějaká souvislost (korelace) mezi procentem otevřených slabik na konci verše a v cézuře. Existence korelačního vztahu se matematicky zjišťuje tak, že se pro každý „vzorek" násobí odchylky od průměrné frekvence otevřených slabik před cézurou a na konci verše a součiny se sčítají: není-li mezi výskytem otevřených slabik na konci a uprostřed verše korelační vztah, jsou-li odchylky náhodné, blíží se součet jejich součinů nule, jednotlivé odchylky se ruší (pro přesné vyčíslení korelace se ovšem užívá složitějšího vzorce):

	Odchylky v cézuře	Odchylky na konci verše	Součin
I	— 4	+1	— 4
II	— 9	—2	+18
III	— 7	+2	—14
IV	+ 6	—6	—36
V	+11	+4	+44
VI	+ 8	0	0
VII	0	—6	0
VIII	—10	+2	—20
IX	+ 7	+3	— 5
X	— 1	+5	— 5
		Součet	—52

Tento výsledek svědčí o slabé disimilační tendenci: zvýšení počtu otevřených slabik na konci 1. půlverše je často doprovázeno snížením jejich

procenta na konci 2. půlverše a naopak. Matematicky zjištěnou korelaci by ovšem bylo třeba interpretovat: a) jaký je její původ, b) jaký je účinek. Původ bude mít tato korelace v silné uspořádanosti větné stavby Holečkova verše: je-li na konci 1. půlverše slovo gramatické kategorie, u níž převažuje zavřená tendence, bývá na konci verše slovní typ s převahou otevřených zakončení a naopak:

Bratr seděl mezi branou,
duhu sobě vystruhoval:
„Pročpak pláčeš, bědná sestro ...“

Protikladu mezi slovesnými tvary a tvary jmennými (zájmennými) je kompozičně využito k pointování půlveršů a jako průvodní jev se přitom objevuje disimilační rozložení otevřených a zavřených slabik. Zdá se, že estetickou funkci protikladu otevřenosti slabik bude třeba zkoumat v souvislosti s dalšími složkami verše a že pro malou výraznost bude obtížné ji určovat izolovaně.

Rým

4.1 Jednotkami inventáře rýmového kódu nejsou slabiky, ale skupiny stejně zakončených slov, tzv. rýmové skupiny, protože to jsou třídy zahrnující prvky o stejném pozičně-kombinatorickém zařazení. Rýmový inventář má proti inventáři zvukosledu daleko větší počet jednotek, a ty mají následkem toho daleko menší frekvenci.[25] Inventář zvukosledu v češtině se skládá z 35 fonémů a jejich průměrná frekvence je asi 3 %. Jen dvoufonémových slabičných kombinací je v češtině teoreticky možných cca 1000 a jejich průměrná frekvence by byla 0,1 %. Kromě toho prvky zvukosledného inventáře jsou identické („všechna *a* jsou stejná“), je irelevantní k jakému slovu patří (samozřejmě pokud situaci nekomplikuje jejich pozice na začátku slova apod. a pokud nejsou součástí slov zvukomalebných apod). Prvky rýmového inventáře nejsou shodné, mají každý svou individuální existenci zdůrazněnou ještě tím, že počet členů některých rýmových skupin je omezený; slova, ba i mnohé slabiky, mají vedle diferenčních příznaků akustických také příznaky sémantické a gramatické, takže jsou současně členy jiných množin než rýmových. Protože jednotky nejsou identické, nestaví schéma rýmového kódu na opakování *shodných prvků,* ale na *analogii* mezi rýmujícími se slabikami. Tato analogie je nejtěsnější u slabik, které jsou v průniku všech zúčastněných množin: akustické, sémantické, gramatické apod. — takový případ jsou gramatické rýmy.

[25] Srov. P. Guiraud, *Problèmes et méthodes de la statisque linguistique,* Paris 1960, 39.

Počet prvků rýmového inventáře je závislý na tom, kolik různých hláskových kombinací se vyskytuje ve slabikách daného jazyka. Např. v angličtině, která má 42 fonémů, je množství teoreticky možných permutací jen dvou fonémů 42^2, tj. asi 1600; ve skutečnosti je rýmových skupin sotva 400. V češtině je z 1000 možných dvoufonémových kombinací 300 neobsazeno,[26] ovšem počet možností zde budou zvyšovat skupiny souhlásek. Neúspornost kódu, pokud jde o jednotky inventáře, vede a) k poměrně častému opakování téhož souzvuku v řadě rýmů (R. F. Brewer vypočítal, že ve 100 rýmových dvojicích se u anglických básníků Popea, Goldsmithe atd.[27] opakuje tentýž souzvuk v 17—43 dvojicích), b) k poměrně snadnému tvoření rýmů (v řadě n slov se často vyskytují slova stejného zakončení).

Bylo matematicky zjišťováno, jak obtížné je tvoření rýmových dvojic v jednotlivých jazycích. A. N. Kolmogorov[28] vypočítal, že v ruštině je pravděpodobnost, že se dvě náhodně vybraná slova budou rýmovat, zhruba 0,005. Protože ze 20 slov je možno sestavit 190 párů, dá se očekávat, že mezi 20 slovy se vyskytne asi 1 dvojice rýmů; obecně:

$$M_2 = \frac{n^2 - n}{2} \cdot 0{,}005 = \frac{n^2 - n}{400}$$

Stanovil také pravděpodobnost pro rýmové trojice (M_3), čtveřice (M_4) atd.

n	Počet párů	M_2	M_3	M_4	M_5	M_{10}
10	45	0,25				
20	190	1	0,03	0,008		
50	1 225	6,25	0,52	0,032		
100	4 950	25	4,17	0,52	0,052	
200	19 900	100	33,3	8,3	1,67	0,00022
500	124 750	625	520	325	162	7

Tyto údaje dovolují odhadnout poměrnou obtížnost rýmových schémat s několikanásobným opakováním rýmů. Např. k vytvoření 7 rýmových dvojic by stačilo asi 60 *různých* slov, což je jen několikrát více, než kolik jich obsahuje 14 veršů sonetu. Pro 3 čtyřnásobné rýmy by již bylo zapotřebí asi 350 různých slov, a takovou slovní zásobu můžeme předpokládat teprve v dosti dlouhé skladbě. (Ukazuje se ostatně také, že opakování rýmů, jež bylo zjištěno u anglických básníků 18. století, nebude patrně jen věcí náhody, ale stylizace.)

Pro francouzštinu vypočítal obdobně P. Guiraud[29], že n slov by mělo teoreticky dát $37 \cdot \left(\frac{n}{100}\right)^2$ rýmů. Je pozoruhodné, že jeho údaje se velmi blíží Kolmogorovým údajům pro ruské rýmy:

[26] Je to patrné z Doleželovy tabulky v cit. článku.
[27] R. F. Brewer, cit. kn., 152.
[28] *Strukturno-typologičeskije issledovanija,* Moskva 1962, 288.
[29] P. Guiraud, *Langage et versification,* 108—109.

Počet slov	Počet rýmových dvojic	
	ruština	francouzština
10	0,25	0,37
100	25	37
500	625	925

Propočítávání rýmových možností jednotlivých jazyků nebude bez významu pro srovnávací teorii rýmu.

4.2 Z hlediska očekávanosti je rýmové slovo v protikladné situaci: čím více členů má rýmová skupina, tím menší prediktabilitu v ní má každé jednotlivé slovo — současně však čím je skupina početnější, zhruba tím častěji se bude její souzvuk objevovat a tím bude mít větší prodiktabilitu, tedy rýmové slovo z početnější rýmové skupiny bude málo překvapivé, pokud jde o zvukovou stránku, ale hodně překvapivé, pokud jde o stránku významovou, tj. pokud vystupuje jako lexikální jednotka. Větší či menší *překvapivost* spojení dvou slov do rýmové dvojice se dá vyjádřit *pravděpodobnostmi přechodu.* Ve skupině monosylab na *-en* mají jednoslabičná slova podle *Frekvence českých slov* ... tuto průměrnou frekvencii: jen ($6477 . 1{,}6 . 10^{-6}$), den ($2287 . 1{,}6 . 10^{-6}$), sen (731 ...), kmen (143 ...), ven (187 ...), len (12 ...), plen (5 ...), sten (4 ...), klen, křen atd. Nebereme-li v úvahu významové spoje, dobový styl atd., pak pravděpodobnost výskytu rýmové dvojice kmen — len $= 143 . 1{,}6 . 10^{-6} . 12 . 1{,}6 . 10^{-6}$, pravděpodobnost výskytu dvojice plen — sten $= 5 . 1{,}6 . 10^{-6} . 4 . 1{,}6 . 10^{-6}$, je tedy 86 krát menší (tato čísla se podstatně nezmění, ani vezmeme-li v úvahu výskyt těchto slov jen v poezii).

Rytmus

5.1 Inventář prvků metrického schématu je velmi prostý: 1. slabika přízvučná, 2. slabika nepřízvučná. Jde tedy o binární kód, jehož entropie je nejvyšší při pravidelné alternaci obou jednotek, tj. v jambu nebo trocheji (H max. $= 1$); v české próze je frekvence obou prvků 0,37 a 0,63, a tedy $H_0 = 0{,}93$ v trojdobých metrech je frekvence 0,33 a 0,63 a $H_0 = 0{,}92$ (je tedy bližší próze než u dvoudobých meter).

Právě proto, že rytmické prvky mají jasně diskrétní charakter a jsou snadno počítatelné, má v této oblasti statistický rozbor dlouhou tradici. Je však třeba říci několik poznámek ke vžité metodě rytmických statistik, a to a) k přesnosti jejích výsledků, b) k obvyklému statistickému postupu.

Při interpretaci rytmických statistik jde o to, do jakých detailů je možno statistiku interpretovat, tj. které rozdíly ve frekvenci přízvuků na jednotlivých slabikách máme pokládat ještě za aleatorické výkyvy a kdy již svědčí

o stylistické tendenci. K tomu musíme znát pro jednotlivé údaje aspoň teoretickou hodnotu průměrné statistické chyby, již vypočteme podle vzorce

$$e = \sqrt{\frac{pq}{n}}$$

přičemž $p = \%$ výskytu, $q = 100 - p$, $n =$ počet analyzovaných veršů. Pro potřeby rytmických statistik předkládám tabulku průměrné statistické chyby (v %) pro jednotlivé desítky procent a pro 100, 200 a 1000 analyzovaných veršů:

%	10	20	30	40	50	60	70	80	90
100	3	4	4,6	4,9	5	4,9	4,6	4	3
200	2,1	2,8	3,2	3,5	3,5	3,5	3,2	2,8	2,1
1000	1	1,3	1,4	1,5	1,6	1,5	1,4	1,3	1

To znamená, že při interpretaci musíme brát v úvahu hodnotu $p \pm e$, tj. pro excerpci 100 veršů nemůžeme z rozdílu 45 % : 55 % činit závěry, protože tyto hodnoty se ještě pohybují v oblasti aleatorických výkyvů kolem hodnoty 50 % (50 ± 5). Statistické chyby jsou největší kolem frekvence 50 % a poněkud klesají u menších a větších hodnot — ovšem s výjimkou velmi nízkých hodnot, pro něž již neplatí rozdělení Gaussovo, ale Poissonovo[30]. Průměrná statistická chyba se v moderních statistických rozborech snižuje tím, že se místo jednoho delšího úseku excerpuje několik kratších vzorků a průměrné hodnoty se vypočítají s přihlédnutím k rozptylu statistických výsledků.

5.2 Rozložení přízvuků ve verši je možno statisticky analyzovat dvěma směry:

a) vertikálně — ve verších se prostě ve vertikálních sloupcích spočítají přízvučné slabiky a statistické výsledky ukáží, jaká je poměrná frekvence přízvuků na 1 ... *n*-té slabice verše, a jak důsledně je tedy v jednotlivých bodech verše realizováno metrum;

b) horizontálně — zjistí se frekvence 1. typu se slovními přízvuky na všech metricky těžkých dobách, 2. typu s „vynechaným" přízvukem na prvním iktu atd.

Vertikální statistika je výhodná pro verš izosylabický a v naší moderní metrice převládá; může ovšem dát zkreslené výsledky, sčítáme-li komplementární rytmické typy — např. statistika strofy, v níž by se pravidelně střídal jambický a trochejský verš, by nevykazovala žádnou rytmickou tendenci. Horizontální statistika je výhodná pro verše s rozkolísaným syla-

[30] Srov. P. Guiraud, *Problèmes ... passin.*

bismem, proto jí běžně užívá např. anglická a ruská metrika. Vedle těchto praktických zřetelů se v posledních letech začíná dávat přednost horizontální statistice také z teoretických důvodů; respektuje lineární povahu verše, tj. dovoluje lépe zkoumat rozloženi jednotek v řadě a uspořádání řad za sebou.

Možnosti obou typů statistiky předvedu na osmislabičném trocheji Holečkovy *Kalevaly*. Vertikální statistika přízvuků:

1	2	3	4	5	6	7	8
100	0	57	0	98	0	38	0

Z výšky vrcholů můžeme soudit na tendenci k dvoudílnosti (tj. k cézuře po 4. slabice), resp. na silnou tendenci k sestupným dipodiím. Další detaily rytmického uspořádání ukáže matrice rytmických různotvarů:

	I	II	III	IV
Čtyřvrcholový v.				
I, II, III, IV	14	14	14	14
Třívrcholový v.				
I, II, III	41	41	41	—
I, II, IV	2	2	—	2
I, III, IV	22	—	22	22
Dvouvrcholový v.				
I, III	21	—	21	—
Součet	100	57	98	38

Matrice rytmických typů ukazuje, že dominantní je verš třívrcholový (65 %), daleko řidší je typ dvouvrcholový (21 %) a čtyřvrcholový (14 %), jednovrcholový se nevyskytuje. Dále je možno přesněji charakterizovat tendenci k dipodičnosti: asi polovina půlveršů je organizována dipodicky (má jen 1 přízvuk), a to téměř výhradně do sestupných dipodií; tato dipodičnost je zvláštně silná v 2. půlverši (63 %).

Matrice rytmických typů má dále tu výhodu, že se rytmické jednotky neabstrahují do jediného statistického modelu, ale pracuje se s celými lineárními strukturami, příp. i s jejich sledem, a je možno zjišťovat a) jaký mají jednotlivé různotvary verše rytmický charakter, b) zda je jejich střídání náhodné nebo kompozičně úkonné. Srovnejme charakter 3 dominantních typů v *Kalevale:*

A. Třívrcholový v. I, II, III

Schránu písní otevříti,
tlumok zpěvu rozvázati,
klubko věšteb rozvinouti,
uzel kouzel rozmotati.

B. Třívrcholový v. I, III, IV

za stádem se brával skotu,
na medná jej honil luka...
za Muurikki chodě černou,
v společnosti straky Kimmo.

C. Dvouvrcholový v. I, III

se křovisek naškubal jsem,
s letorostů nalámal jsem,
na travinách nasbíral jsem,
na pěšinách posbíral jsem.

Již z uvedených příkladů je zřejmé, že jednotlivé typy mají tendenci se seskupovat; je to přirozený důsledek stylistických paralelismů, ale současně se tím tyto paralelismy zvýrazňují. Všechny tři typy verše jsou zřetelně dvoudílné, přičemž u typu A je významově zatíženější 1. půlverš, u typu B 2. půlverš, a typ C dodává stylu epickou šíři, uvolňuje tok myšlenky, protože v každém půlverši je jen jedno významové jádro.

Srovnejme s Holečkovým trochejem čtyřstopý trochej Nerudův (*Balada zimní* a *Romance štědrovečerní*):

	I	II	III	IV
Čtyřvrcholový v.				
I, II, III, IV	49	49	49	49
Třívrcholový v.				
I, II, III	18	18	18	—
I, II, IV	7	7	—	7
I, III, IV	17	—	17	17
Dvouvrcholový v.				
I, III	8	—	8	—
I, IV	1	—	—	1
Součet	100	74	92	74

Vertikální součet prozrazuje, že Nerudův verš má proti verši Holečkovu výraznější rytmus a rytmicky obdobně řešené oba půlverše. Matrice dovoluje tento poznatek konkretizovat: výrazně dominantní je u Nerudy typ s přízvuky na všech čtyřech těžkých dobách:

Zaklel, máchnul rukou v kole,
hup! a již jsou všichni dole.

Typy s lehčí dobou na první polovici verše jsou zastoupeny stejně často a jsou také rozloženy bez zjevných kompozičních tendencí, takže u Nerudy je kompoziční využití rytmických různotvarů slabší než u Holečka.

5.3 Srovnáním frekvenčních matric rytmických různotvarů pro různe autory je možno stanovit obecné tendence, *invarianty*, dané české formy a konfrontovat je s invariantními rysy téže formy v jiných jazycích. Kon-

frontujeme-li údaje získané pro český osmislabičný trochej s údaji pro osmislabičný ruský trochej, jak je zpracoval G. Šengeli[31], zjistíme:

a) ze 16 teoreticky možných forem se v praxi užívá asi 5–6; nultou nebo mizivou frekvenci mají formy s jedním přízvukem (4 možnosti), forma bez jediného přízvuku (1 možnost), dvoupřízvukové formy, v nichž oba přízvukové vrcholy souvisí (3 možnosti). To jsou invariantní vlastnosti osmislabičného trocheje českého i ruského.

b) Diferenční rysy české a ruské rytmické řady se objeví, postavíme-li proti sobě na jedné straně střední výskyt jednotlivých variant u Holečka a Nerudy, na druhé straně střední hodnoty udávané pro ruský čtyřstopý trochej Šengelim; v pravé polovině tabulky označíme frekvenci dosahující aspoň 10 % znakem +, frekvenci blízkou nule znakem —:

	Č	R	Č	R
I, II, III, IV	32	14	+	+
I, II, III	30	0	+	—
II, III, IV	0	21	—	+
I, II, IV	5	35	—	+
I, III, IV	20	0	+	—
I, III	10	0	+	—
II, IV	0	19	—	+

Konfrontujeme-li poměrnou frekvenci jednotlivých různotvarů v českém a ruském verši, vidíme, že — až na první typ — jsou frekvence jednotlivých různotvarů českých a ruských v poměru nepřímé úměrnosti: v českém verši převažují formy, v nichž jsou přízvuky soustředěny v první části verše (půlverše). To je invariantní rys českého (ruského) inventáře rytmických různotvarů, pokud jde o jejich frekvenci. Příčinou je tendence k sestupným kólům v češtině a ke vzestupným kólům v ruštině.

5.4 Při analýze vyšších celků (strof, celé básně apod.) je možno pracovat s rytmickými různotvary jako s *jednotkami* a zjišťovat a) v jaké míře bylo využito jejich střídání pro rozrůznění rytmického pohybu skladby, b) zda jsou rozloženy náhodně, příp. nápadně nerovnoměrně, a zda jsou tyto výkyvy v díle funkční. Při této analýze je možno užít hodnoty *entropie* jako pomocné míry, která dovoluje vyjádřit proměnlivost řady veršů v několika skladbách příp. jazycích (např. v básni psané čtyřstopým trochejem $H\max. = 4$, H_1 pro Holečkovu *Kalevalu* = 1,99, u Nerudy $H_1 = 2{,}01$ — tedy mezi oběma autory není po této stránce podstatný rozdíl, oba poměrně málo využili kombinačních možností). Zjišťování rozptylu v rozložení je ovšem velmi pracné a jen v některých případech budou jeho výsledky použitelné jako pomocné údaje pro literární analýzu.

[31] G. Šengeli, *Technika sticha*, Moskva 1960, 174 n.

6.1 Pokusy o vyčíslení inventáře sémantických jednotek se zatím pohybují v úrovni hypotéz. Prvním spolehlivějším krokem ke kvantitativní analýze uspořádání významů v jazykovém sdělení by bylo zjištění, jak vypadá distribuce slov podle gramatických kategorií. Pro zkoumání, jakým způsobem ovlivňuje veršové schéma uspořádání slovních kategorií v průběhu verše, a tím veršovou řadu významově struktuuje, je třeba vyčíslit, jak často jsou 1 . . . n-tá slabika verše obsazeny jednotlivými slovními kategoriemi. Z těchto údajů je pak možno také vypočítat průběh entropie gramatického skladu verše; čím je entropie vyšší, tím proměnlivěji je využito všech kategorií, čím je nižší, tím je rozložení nerovnoměrnější, „stylizovanější". Uvedu údaje pro 3 typy verše (v %):

A. Osmislabičný nerýmovaný trochej s cézurou (Holeček, *Kalevala*)

	I	II	III	IV	V	VI	VII	VIII
I. subst.	32	44	66	60	29	34	31	32
II. slov.	10	11	14	9	48	47	58	55
III. adj.	9	11	6	4	7	7	9	8
IV. adv.	11	8	5	4	2	3	1	1
V. zájm.	3	14	6	21	7	8	1	4
VI. předl.	25	3	2	1	6	1	—	—
VII. spoj.	10	9	1	1	1	—	—	—
VIII. čísl.	—	—	—	—	—	—	—	—
IX. citosl.	—	—	—	—	—	—	—	—
H_1	2,51	2,38	1,69	1,73	1,99	2,10	1,42	1,55

B. Blankvers (Zeyer, *Karolinská epopeja*)

	I	II	III	IV	V	VI	VII	VIII	IX	X
I	10	35	37	33	24	27	24	26	32	30
II	10	21	21	30	22	33	30	34	34	28
III	—	12	13	10	10	12	13	23	25	20
IV	16	7	10	9	7	10	15	2	4	8
V	19	7	11	7	18	10	10	2	4	12
VI	12	11	2	10	7	6	1	13	1	2
VII	33	4	3	1	12	1	6	—	—	—
VIII	—	3	3	—	—	1	1	—	—	—
IX	—	—	—	—	—	—	—	—	—	—
H_1	2,44	2,60	2,50	2,46	2,37	2,45	2,52	2,14	1,99	2,27

C. Jedenástslabičný rýmovaný jamb (Vrchlický, *Betlem*)

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI
I	12	49	53	44	33	38	34	36	35	59	59
II	1	17	19	20	18	18	18	16	13	15	15
III	4	12	13	14	12	21	21	22	20	19	18
IV	22	8	7	9	15	10	9	7	8	3	5
V	20	4	4	3	15	1	12	4	8	4	5
VI	20	10	—	9	3	10	3	14	8	—	—
VII	18	—	3	1	3	3	3	1	8	—	—
VIII	—	—	—	—	1	—	—	—	—	—	—
IX	3	—	—	—	—	—	—	—	—	—	—
H_1	2,65	2,12	1,95	2,23	2,55	2,32	2,43	2,36	2,56	1,66	1,67

Tabulky distribuce slovních druhů ve verši dovolují formulovat některé prozatímní závěry:

a) Některé slovní druhy jsou ve verši rozloženy poměrně rovnoměrně a nezávisle na průběhu verše: adverbia, číslovky. Některé slovní druhy se velmi zřetelně hromadí v určitých bodech verše a půlverše (zpravidla okrajových), nebo se jim vyhýbají: substantiva, slovesa. Konečně některé druhy vykazují rozdílné tendence v různých stylistických typech: adjektiva.

b) Substantiva tíhnou spíše ke středu a konci verše (půlverše), zřídka podstatným jménem verš (zvláště jambický) začíná.

Slovesa se také zřídka objevují na začátku verše, v některých textech je jejich frekvence vyšší v 2. půlverši (Holeček, Zeyer), což svědčí o tendenci umísťovat v 1. půlverši subjekt a v 2. půlverši predikát.

Přídavná jména u Vrchlického a Zeyera tíhnou spíše ke konci verše, u Holečka na počátek verše (je to důsledek archaizovaného epického stylu i trochejského rytmu).

Příslovce se hromadí na počátku řady a jinak jsou rozložena poměrně rovnoměrně.

Zájmena se výrazně seskupují na počátku veršů (půlveršů) u Vrchlického a Zeyera (1. a 5. slabika), u Holečka na konci 1. půlverše (4. slabika); rozdíly jsou zde způsobeny také tím, že nejsou odlišena osobní zájmena, častá na počátku syntagmat, a zájmena přivlastňovací, která se inverzí často dostávají na konec syntagmat (Holeček).

Předložky se obyčejně — ale ne vždy — hromadí na počátku rytmických úseků; protože jsou v češtině vždy přízvučné, hromadí se na metricky důrazných slabikách.

Spojky se velmi výrazně kupí na počátku verše, někdy i na počátku 2. půlverše (což svědčí o syntaktické dvoudílnosti veršů).

Číslovky a citoslovce jsou vzácné a vyskytují se spíše na počátku veršů. Vcelku lze tedy uzavřít, že jsou dva typy slovních druhů: podstatná jména, slovesa a přídavná jména se hromadí spíše uprostřed a ke konci rytmických celků a tvoří jejich významové jádro. Ostatní slovní druhy, které mají funkci především syntaktickou, vztahovou se kupí na počátku veršů (zvláště jambických).

6.2 Pro přesné a jednoduché vyjádření míry diferenciace verše, pokud jde o slovní druhy, je užitečné vypočítat entropii rozložení slovních druhů pro jednotlivé slabiky (příslušné hodnoty jsou uvedeny v posledním řádku tabulek). Průběh entropie ve verši dovoluje činit tyto závěry:

a) Hodnota entropie nápadně klesá na konci verše, především na předposlední slabice (na poslední bývá obyčejně slabý vzestup podmíněný poměrně velkou frekvencí jednoslabičných zájmen). Pokles vzniká tím, že ke konci verše se vyskytují prakticky jen podstatná jména, přídavná jména a slovesa, ostatní slovní druhy buď vůbec ne, nebo velmi zřídka; je zde tedy nejstereotypnější slovní sklad. Totéž platí o konci půlverše.

b) Vysoká je po této stránce entropie na počátku verše, zvláště začínají-li tam často syntaktické celky (Holeček, Vrchlický); zde se totiž nejčastěji vyskytují řídké slovní druhy (spojky, předložky, zájmena, adverbia), slovní sklad je zde nejrozrůzněnější.

*

Poetika a matematika se v poslední době někdy sbližují do té míry, že z toho vznikají specifické obtíže a nebezpečí. Domnívám se, že procházíme dosti nevýhodnou a nepopulární etapou ve vývoji metod technické analýzy literárního díla. Stává se zřejmým, že nevystačíme jen s přibližnými a subjektivními soudy o literatuře a že máme možnost si přesně vypočítat některé síly, které působí uvnitř literárního díla. Zdá se, že se část filologických pracovníků bude muset zabývat matematickou analýzou, podobně jako se třeba v době humanismu museli zabývat pracným sestavováním slovníků. Toto přechodné stadium sestavování matematických programů má svá nebezpečí:

a) někteří estetikové vidí v matematice nikoliv pomocnou vědu, ale východisko všech estetických kategorií, a neuvědomují si, že při čtenářském vnímání se vztahy mezi některými složkami díla konkretizují jinak, než by odpovídalo jejich kvantitativnímu uspořádání;

b) nutnost vypočítat některé matematické veličiny, jako je entropie, vede k fetišizaci těchto veličin, chápou se jako cíl a ne jako nástroj. Základním faktem zůstává, že cílem literárněvědné práce není poznávat matematické zákonitosti, ale užívat matematických zjištění jako pomocných údajů k interpretaci literatury, stále se ptát po estetickém smyslu s relevancí získaných kvantitativních údajů.

PRELIMINARIES TO AN INFORMATION ANALYSIS OF VERSE

1. Versification is regarded as code, i. e. as a system of schemes regulating the grouping of discreet elements into units of linear structure, subject to probability laws. This permits to use stochastic mathematics in analysing certain features of verse.

2. Entropy of sounds in connected texts of various stylistic types is evidence of the fact that poetry prefers the frequent sounds and an equal distribution of sounds throughout the text. Even chance grouping of sounds may be endowed with an aesthetic function.

3. The distribution of open v. closed syllables seems to be more dependent on syntactic limits than on verse limits and to acquire an aesthetic function in connection with other stylistic elements only.

4. New possibilities of a comparative analysis of rhyme are given by the computation of rhyming facilities in different languages and of transitional probabilities between pairs of rhyming words.

5. „Horizontal" statistics of rhythmical variants of a measure offers the advantage of preserving the linear character of the units; moreover, they permit to compute the entropy, i. e. the degree of ordered arrangement, of greater wholes, where the single lines act as units.

6. As a first step towards an analysis of the semantic structure of verse the distribution of the 9 grammatical categories on the 1 ... n-th syllables of verse has been computed. The profile of entropy bears evidence of the fact that the point of greatest differentiation in grammatical categories is the beginning of the line, the point of least differentiation the end of line.