

ENTROPIA A SÚČASNÁ SLOVENSKÁ LITERÁRNA TVORBA

Marianna KRAVIAROVÁ

Lingvistiku je možné skúmať aj exaktnými vednými disciplínami a tým poukázať na jej kvantitatívny rozmer, na vzťahy medzi javmi v danom systéme. Jednou z možností je využitie poznatkov teórie informácie aj v literatúre.

V roku 1948 vznikla nová vedná disciplína – kybernetika, ktorá sa zaoberá riadením a komunikáciou akéhokoľvek systému, ktorej zakladateľom je americký matematik Norbert Wiener. Kybernetiku delíme na teoretickú a aplikovanú, podľa oblastí ktorým sa venuje. K teoretických odborov patrí aj teória informácie zaoberajúca sa získavaním, prenosom, spracovaním, ukladaním a využívaním informácií, ktorej autorom je Claude Elwood Shannon. Základy tomuto vednému odboru položil v roku 1948 prácou *Matematická teória komunikácie*. Informácie skúma v rôznych druhoch systémov – ekonomickom, spoločenskom, biologickom, či technickom. Na tomto mieste je vhodné vysvetliť pojem systém. Jedna z mnohých definícií systému reprezentuje systém ako množinu prvkov a vzťahov medzi nimi. Z tohto princípu vychádza aj využitie teórie informácie v takých vedných disciplínach akými sú jazykoveda, či literárna veda. Teória informácie využíva najmä analytické a pravdepodobnostné metódy. K základným pojmom teórie informácie patrí entropia, ktorú je možné klasifikovať ako mieru neurčitosti, neusporiadanosti, dezorganizácie systému. Vzorec na výpočet reálnej entropie je

$$H = - \sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

H – reálna (absolútna) entropia meraná v bitoch,

p – pravdepodobnosť výskytu daného javu (prvku) v systéme,

n – počet skúmaných javov (prvkov) v systéme, rozsah súboru.

Pravdepodobnosť výskytu javov v systéme sa môže pohybovať v rozpätí $p \in \langle 0; 1 \rangle$, pričom platí, ak je pravdepodobnosť nulová, t. j. $p = 0$, daný jav nemôže v systéme nastať. Druhá krajná

poloha rozpätia pravdepodobností $\mathbf{p} = \mathbf{1}$ znamená, že udalosť, jav nastane v každom prípade. Súčet všetkých pravdepodobností skúmaného súboru je rovný jednej, t. j. $p_1 + p_2 + \dots + p_n = 1$, alebo je možné ho vyjadriť v percentách, ak pravdepodobnosť

$$\sum_{i=1}^n p_i = 100\%$$

vynásobíme číslom 100, potom je $\sum_{i=1}^n p_i = 100\%$. Reálna entropia je rovná súčtu súčinov pravdepodobností jednotlivých skúmaných javov systému a dvojkových logaritmov pravdepodobností týchto javov vynásobená číslom -1, pretože dvojkové logaritmy čísel z rozsahu $\langle 0; 1 \rangle$, ktoré zodpovedajú hodnotám pravdepodobností, sú záporné hodnoty a absolútna entropia je číslo kladné, ktoré vyjadruje akú informáciu systém prenáša. Vzťah medzi entropiou a informáciou je vzťahom nepriamo úmerným. Ak o systéme získavame informácie, zvyšuje sa jeho miera informácie, usporiadanosti, určitosti, organizácie a zároveň sa znižuje miera entropie, neusporiadanosti, dezorganizácie systému. Množstvo prenášanej informácie určitou jazykovou jednotkou je tým väčšie, čím menšia je pravdepodobnosť jej výskytu a teda aj frekvencia tejto jednotky. Na základe týchto poznatkov informáciu nazývame aj negatívnu entropiou, resp. negentropiou. Reálna entropia môže nadobúdať dve krajné polohy. Ak je pravdepodobnosť jedného prvku systému rovná jednej a ostatné sú nulové, reálna entropia nadobudne nulovú hodnotu, t. j. $H = 0$. Ak všetky javy v systéme

môžu nastať s rovnakou pravdepodobnosťou $p_i = \frac{1}{n}$; pričom n je počet rôznych stavov v systéme, je absolútna (reálna) entropia maximálna, miera neurčitosti systému je maximálna a matematicky to vyjadríme vzťahom

$$H_{max} = \log_2 n \quad (2)$$

C. E. Shannon (1948) okrem reálnej entropie H uvádza aj relatívnu entropiu h , ktorá je daná vzťahom

$$h = \frac{H}{H_{max}} = \frac{-\sum_{i=1}^n p_i \log_2 p_i}{\log_2 n} \quad (3)$$

Relatívna entropia je pomer reálnej entropie k entropii maximálnej, ktorá môže v danom systéme nastať. Relatívna entropia sa využíva na porovnávanie systémov alebo súborov, ktoré nemajú

rovnaký počet prvkov, tzn. n nadobúda rôznu hodnotu pre rôzne systémy. S entropiou úzko súvisí aj výpočet redundancie, pre ktorú platí

$$R = 1 - h = 1 - \frac{H}{H_{\max}} = 1 - \frac{-\sum_{i=1}^m p_i \log_2 p_i}{\log_2 n} \quad (4)$$

Redundancia môže nadobúdať hodnoty $\langle 0; 1 \rangle$. Redundancia systému znamená nadbytočnosť, ale nie zbytočnosť prvkov v systéme. Čím je vyššia frekvencia istého prvku v systéme, tým je vyššia aj jeho pravdepodobnosť a teda aj redundancia, a tým menšie je množstvo informácie, ktorú prenáša. Redundancia zaisťuje aj spoľahlivosť prenosu informácií. Význam redundancie je v tom, že pri prenose informácií informačným kanálom dochádza často aj k určitým poruchám, ktoré sú spôsobené rôznymi faktormi ako nesprávna výslovnosť hovoriaceho, chyby tlače, nepozornosť poslucháčov a pod. Tieto poruchy označujeme spoločne pojmom šum. Ak by sa v jazyku dodržiavalo iba ekonomické hľadisko a žiadna redundancia, aj pri najmenšej poruche v prenosovom kanáli, by došlo k znehodnoteniu informácie v takom rozsahu, že by sa stala nezrozumiteľnou (J. Černý, 1996, s. 277). Výpočet entropie, či redundancie je možné aplikovať na akékoľvek javy, ktoré využívajú svoju pravdepodobnosť výskytu. Pri aplikovaní entropie a redundancie v jazykovednom a literárnovednom výskume je nutné najskôr skúmaný jav vymedziť (Sabol - Zimmermann, 1986, s. 70). Vplyv entropie na systém, v našom prípade na literárne dielo, môžeme pozorovať zmenou jej hodnoty, pričom nižšia entropia ukazuje na tendenciu k ustálenosti systému, vyššia hodnota entropie poukazuje na uvoľnenosť vzťahov medzi prvkami systému, na určitý stupeň neusporiadanosti systému (Kráľ - Sabol, 1989, s. 348).

Na skúmanie vplyvu veličín teórie informácie v súčasnej slovenskej literatúre sme vybrali štyri typy textov (súborov, systémov) - verš viazaný, verš voľný, prózu a odborný text. Ide o pôvodné diela súčasných slovenských autorov.

Vychádzajúc z konkrétnych aplikácií veličín teórie informácií vo výskumoch umeleckých textov (Sabol, 1973, Sabol - Zimmermann, 1986) sme skúmali ako prvé frekvenciu výskytu sonantických jadier slabík a ich vplyv na výstavbu textov básní a prózy.

Z výskumov, ktoré boli realizované v 60-tych až 80-tych rokoch 20. storočia vyplývajú závery, že výskyt samohlások v poézii je rovnomernejší ako v próze. Pokúsime sa tieto závery potvrdiť aj v tvorbe súčasných slovenských autorov viazaných a voľných veršov, umeleckej prózy a odborných textov. Vybrali sme trinásť básní P. Garana z jeho nového rukopisu *Pozemské triolety*, ktorá využíva výstavbu viazaného verša, jedenásť básní M. Milčáka zo zbierky básní *Siedma kniha spánku* (voľný verš), úryvok z prózy T. Keleovej-Vasilkovej *Srdce v tme* (s. 100 – 102) a úryvok z odborného textu J. Dolníka *Všeobecná jazykoveda* (s. 20 – 21). Z básnických zbierok sme vybrali väčšie množstvo básní, aby sme získali dostatočné množstvo skúmaných prvkov. Pri malých súboroch sa môžu hodnoty entropie a redundancie odlišovať od hodnôt zistených pri väčších súboroch, čo je dané samotnou výstavbou a funkciou jednotlivých básní. Sonanty sú nositeľmi všetkých suprasegmentálnych vlastností a podieľajú sa aj na eufonickej výstavbe diela, pre jednoduchšie spracovanie dát sme vytvárali vokalické typy spájaním sonánt rovnakých timbrov (Sabol – Zimmermann, 1986, s. 63). Počty sonánt v jednotlivých skúmaných systémoch sa nachádzajú v tabuľke 1.

sonanty	viazaný verš	voľný verš	próza	odborný text
A	205	464	448	336
E	184	383	340	329
I	142	336	287	292
O	162	384	342	329
U	99	157	99	156
Ä	1	4	1	1
R	12	13	6	13
L	0	11	6	0
Počet sonánt	805	1752	1529	1456

Tab. 1. Počty sonantických jadier a celkový počet sonánt v jednotlivých druhoch skúmaných textov.

Pomocou vyššie uvedených vzorcov (1) až (4) sme vypočítali reálnu, relatívnu entropiu a redundanciu sonánt jednotlivých skúmaných diel. Vypočítané hodnoty H , h , R sa nachádzajú v tabuľke 2.

	viazaný verš	voľný verš	próza	odborný text
reálna entropia (H)	2,371	2,354	2,265	2,336
relatívna entropia (h)	0,844	0,785	0,715	0,832
redundancia (R)	0,156	0,215	0,285	0,168

Tab. 2. Hodnoty reálnej, relatívnej entropie a redundancie sonánt pre štyri druhy textov súčasných slovenských autorov.

Hodnoty relatívnej entropie klesajú od viazaného verša, cez voľný verš až po prózu a poukazujú na skutočnosť, že samohlásky sa využívajú pri básňach písaných viazaným veršom rovnomernejšie ako voľným veršom, čo je dané nízkou hodnotou redundancie, ktorá smerom od poézie k próze narastá. Ešte nižšiu hodnotu entropie pozorujeme pri prozaickom texte. Hodnoty entropie v odbornom texte sú vyššie ako by sme predpokladali, čo môže byť dané zvoleným úryvkom textu. Ďalšia veličina, na ktorej môžeme pozorovať rozdiel medzi poetickým a prozaickým textom na úrovni foném je nižšia samohláskovitost' v poézii. Vo viazanom verši sme skúmali 793 samohlások a 1093 spoluhlások, vo voľnom verši 1728 samohlások a 2421 spoluhlások, v próze 1517 samohlások a 1961 spoluhlások a skúmanému odbornému textu zodpovedalo 1442 samohlások a 1932 spoluhlások. Pomer samohlások a spoluhlások v jednotlivých textoch, ktorý sa premietol do hodnôt entropií a redundancií je zosumarizovaný v tabuľke 3.

	viazaný verš	voľný verš	próza	odborný text
pomer samohlások	0,420	0,420	0,436	0,428
pomer spoluhlások	0,580	0,580	0,564	0,572
reálna entropia (H)	0,981	0,981	0,988	0,985
relatívna entropia (h)	0,981	0,981	0,988	0,985
redundancia (R)	0,019	0,019	0,012	0,015

Tab. 3. Hodnoty pomerov samohlások a spoluhlások, ich reálnej a relatívnej entropie a redundancie pre štyri druhy textov súčasných slovenských autorov.

Hodnota reálnej a relatívnej entropie je totožná, pretože v systéme sa nachádzajú iba dva rôzne prvky (samohlásky a spoluhlásky) a maximálna entropia, podľa vzorca (2), v tomto prípade je rovná hodnote jedna, takže zo vzorca (3) vyplýva, že reálna entropia je rovná entropii relatívnej. Z tabuľky 3 vyplýva, že samohláskovitost' stúpa smerom od viazaného verša (v našom prípade sú zís-

kané hodnoty pre viazaný a voľný verš totožné), cez voľný verš k prozaickému textu. Hodnota relatívnej entropie pre odborný text je o niečo nižšia ako pri próze, ale tendencia stúpania samohláskovitosti od poézie k próze je zachovaná. Zároveň týmto smerom klesá redundancia. Tento záver súvisí aj so zákonitosťou, že dĺžka slabík v slovách stúpa smerom od viazaného verša k próze, a tým stúpa aj množstvo samohlások v texte (Sabol - Zimmermann, 1986, s. 95). Preto sme ďalej skúmali slabikovú dĺžku slov a jej vplyv na rytmickú výstavbu textov. Rozdiel medzi poetickými a prozaickými dielami môžeme pozorovať rôznym stupňom rytmickosti výstavby textu. Tá je daná aj priemernou slabikovou dĺžkou slov (Sabol, 1973, s. 13), pričom rytmicky výhodnejšie sú tie texty, ktoré majú nižšiu slabikovú dĺžku slov čo sa premieťa aj do hodnôt entropie a redundancie, tzn. rytmicky vhodné sú tie systémy (v našom prípade texty), ktoré majú nižšie hodnoty entropie a zároveň vyššie hodnoty redundancie.

	viazaný verš	voľný verš	próza	odborný text
priemerná slabiková dĺžka slov	1,96	2,13	2,04	2,48

Tab. 4. Priemerná slabiková dĺžka slov pre štyri druhy textov súčasných slovenských autorov.

Ak sa pozrieme na nami vybrané skúmané texty (tab. 4) zistíme, že v rukopise P. Garana, ktorý píše viazaným veršom je priemerná dĺžka slabík v slove 1,96. Ide o hodnotu, ktorá je porovnateľná s hodnotami uvádzanými v odbornej literatúre pre poetické texty (Sabol, 1978). Priemerná slabiková dĺžka slov v zbierke básni M. Milčáka zodpovedá skôr prozaickému textu ako poetickému. Najvyššie hodnoty ukazuje odborný text. Počty slabík v slovách skúmaných súborov sa nachádzajú v tabuľke 5.

	viazaný verš	voľný verš	próza	odborný text
1-slabičné slová	151	254	287	142
2-slabičné slová	157	304	237	166
3-slabičné slová	61	179	140	152
4-slabičné slová	33	53	67	77
5-slabičné slová	1	16	10	26
6-slabičné slová	1	6	3	6
7-slabičné slová	1	1	0	1
8-slabičné slová	0	0	0	1

Tab. 5. Slabiková dĺžka slov v skúmaných textoch súčasných slovenských autorov.

S týmto ukazovateľom úzko súvisí aj výpočet entropie a redundancie slov podľa ich slabikovej dĺžky. Vypočítané hodnoty veličín teórie informácie podľa vzorcov (1) až (4) sú uvedené v tabuľke 6.

	viazaný verš	voľný verš	próza	odborný text
reálna entropia (H)	1,810	1,968	1,933	2,218
relatívna entropia (h)	0,700	0,701	0,748	0,736
redundancia (R)	0,300	0,299	0,252	0,264

Tab. 6. Hodnoty reálnej, relatívnej entropie a redundancie slov podľa ich slabikovej dĺžky pre štyri druhy textov súčasných slovenských autorov.

Pretože v skúmaných textoch sme zaznamenali rôznu dĺžku slabík slov (tab. 5), okrem reálnej entropie sme vypočítavali aj hodnoty relatívnej entropie slov podľa ich slabikovej dĺžky, ktoré ukazujú tendenciu nárastu smerom od poézie až k odbornému textu, aj keď nami vybraný úryvok z odborného textu ukazuje nižšiu hodnotu entropie h ako próza, ale tendencia k rytmickosti je zachovaná smerom od prózy až k viazanému veršu.

Uvedené výpočty sú iba základné možnosti využitia poznatkov teórie informácie v jazykovede a literárnej vede, ale ukazujú, že tieto vedné disciplíny je možné skúmať aj exaktnými metódami, ktoré dopĺňajú literárnovedný výskum umeleckých diel. Pri týchto záveroch je potrebné si uvedomiť s akým materiálom pracujeme a ako je dôležité vhodne vymedziť skúmaný jav. Využitím veličín teórie informácie pri skúmaní súčasných slovenských literárnych diel sme potvrdili zákonitosti, ktoré súvisia s výstavbou týchto

textov a to, že počet samohlások v slovách stúpa od viazaných veršov, cez verše písané voľným štýlom až k próze, resp. odbornému textu, čo súvisí s narastajúcim počtom slabík a vyrovnávaním pomeru samohlások a spoluhlások v slovách. Ďalším záverom je skutočnosť, že využívanie samohlások je vo viazaných veršoch rovnomernejšie ako v próze alebo odborných textoch.

Literatúra

- Černý, Jiří.: *Dějiny lingvistiky*. Olomouc: Votobia 1996. 512 s. ISBN 80-85885-96-4.
- Dolník, Juraj: *Všeobecná lingvistika. Opis a vysvetľovanie jazyka*. Bratislava: Veda 2009. 375 s. ISBN 978-80-224-1078-6.
- Garan, Pavol: *Pozemské triolety*. 2011. Z rukopisu.
- Keleová-Vasilková, Táňa: *Srdce v tme*. Bratislava: Ikar 2009. 318 s. ISBN 978-80-551-2011-9.
- Král, Ábel – Sabol, Ján: *Fonetika a fonológia*. Bratislava: Slovenské pedagogické nakladateľstvo 1989, 392 s. ISBN 80-08-00036-8.
- Milčák, Marián: *Siedma kniha spánku*. Levoča: Modrý Peter 2006. 64 s. ISBN 80-85515-69-5.
- Sabol, Ján: *Konsonanticko-vokalické zloženie slovenských slov*. Slovenská reč, 33, č. 5, 1968, s. 281 – 289.
- Sabol, Ján: *Pomer jednotlivých typov konsonantických skupín*. Slovenská reč, 36, č. 2, 1971, s. 71 – 78.
- Sabol, Ján: *Skúmanie a interpretácia rytmických tendencií textu*. Slovenská reč, 38, č. 1, 1973, s. 12 – 22.
- Sabol, Ján: *Z frekvenčného výskumu zvukovej sústavy slovenčiny*. In: *Studia Academica Slovaca*. 7. Prednášky XIV. Letného seminára slovenského jazyka a kultúry. Redaktor Jozef Mistrík. Bratislava: Alfa 1978, s. 499 – 511.
- Sabol, Ján – Zimmermann, Július: *Štatistika. Exaktné metódy v jazykovede a v literárnej vede*. Košice: Rektorát Univerzity P. J. Šafárika 1986. 158 s.
- Shannon, Claude Elwood: *A Mathematical Theory of Communication*. In: *The Bell System Technical Journal*. Vol. 27, s. 379 – 423, 623 – 656, July, October, 1948.
- Štraus, František: *Základy slovenskej verzológie*. Bratislava: Literárne informačné centrum 2003. 251 s. ISBN 80-88878-81-0.