

Polytextia kompozít

Kristína Bobeková

Slovenský národný korpus, Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied,
Bratislava

Pojem polytextia je v slovenskej lingvistike relatívne novým pojmom. Aktuálnosť tohto pojmu potvrdzuje aj skutočnosť, že sa zatiaľ v slovenskej jazykovede nepoužíva, keďže ide o termín prevzatý z nemeckej lingvistiky. Ak už hovoríme o preberaní termínov, musíme spomenúť, že aj metódy matematickej jazykovedy, s ktorými sme pracovali pri skúmaní tejto textovej vlastnosti kompozít, sú prevzaté z nemeckej lingvistiky.

Polytextiu chápeme ako frekvenciu jednej zloženiny v rôznych textoch, resp. kontextoch. Ide o pojem označujúci počet výskytov jazykovej jednotky v textoch celého skúmaného súboru. Kvantitatívne vyjadrenie polytextie nám zaručuje aj extenzívne vyjadrenie ďalších skúmaných vlastností kompozít.

Na lingvistický výskum polytextie a jej vzťahu k iným textovým vlastnostiam, konkrétne k frekvencii, sme využili štatistické metódy a metódy modelovania. Vzťahy medzi textovými vlastnosťami kompozít sme skúmali matematickými a štatistickými metódami formou uplatnenia adekvátnych pravdepodobnostných modelov a znázornením vzťahov medzi jednotlivými vlastnosťami pomocou spojitých funkcií. V tejto etape príspevku sme konzultovali s prof. G. Altmannom¹, ktorý je zakladateľom modernej kvantitatívnej jazykovedy, výber vhodných modelov a výpočtov pomocou softvéru Altmann-Fitter a NLREG, ktoré sme nemali k dispozícii. Pri uplatňovaní štatistických metód ide o zvyčajný postup, že sa výber, definícia a počet získaných dát konzultuje so štatistikom, aby sa údaje dali zmysluplne spracovať. Štatistický softvér Altmann-Fitter filtruje získané údaje vo forme frekvenčnej tabuľky na vyše 200 pravdepodobnostných rozdelení, ktoré sa následne dajú uplatniť na lingvistické údaje, takže konzultácie a výber zodpovedajúcich rozdelení a ich overenie sa zvyčajne realizuje s matematikom. Prof. G. Altmann ako špičkový odborník v oblasti matematickej a synergetickej lingvistiky sa podujal o uvedených bodoch s nami konzultovať a dať získané údaje do fittera na overenie. Konzultácie o matematických metódach nám poslúžili aj na isté overenie postulovaných predpokladov.

¹ Prof. G. Altmann je nemecký lingvista, ktorý mnoho rokov pracoval na Porúrskej univerzite v Bochume. Jeho jazykovedný výskumný záujem sa zakladá na hypotézach o jazykových zákonoch. Tieto sú odvodené z teoretických predpokladov, ktoré sú matematicky formulované a empiricky testované. Príspevok vychádza z diplomovej práce, ktorá vznikla v spolupráci FF UCM v Trnave s prof. G. Altmannom.

Súčasný trend textových vlastností zloženín v slovenskom jazyku sme sa pokúsili zachytiť analýzou publicistických textov, z ktorých sme excerpovali jazykový materiál. Lingvistický výskum sme realizovali na materiáli periodika Plus 7 dní (2011, roč. 21, č. 42 – č. 52), v ktorom sa nachádzajú nielen populárno-vedecké články, ale aj informácie o rôznych spoločenských, kultúrnych, politických i prírodných javoch, ktoré si vyžadujú presné pomenovania. Informačná nasýtenosť sa dosahuje práve kompozitami. Publicistická tvorba je bohatým zdrojom vzniku nových zloženín, ktoré sa dostávajú do praxe a obohacujú slovnú zásobu. Avšak nie každé kompozitum, ktoré sa nachádza v publicistických textoch, sa nachádza aj v jazykovedných príručkách a slovníkoch. Táto skutočnosť je spôsobená aktuálnou potrebou tvorcu nového pomenovania činnosti, resp. javu. A naopak, nie každé kompozitum, ktoré sa využíva v redakčnom úze, je využiteľné, resp. aktívne využívané aj v komunikácii.

Výskumná vzorka jazykového materiálu, s ktorou sme pracovali, obsahuje cca 600 rozličných zloženín, z ktorých sa viaceré opakujú. Môžeme usúdiť, že aj frekvencia kompozít je obrazom súčasného stavu slovenského jazyka. Na potvrdenie tohto synergického efektu by bola potrebná podrobnejšia práca s korpusom a súčasnými údajmi o frekvencii, čo nie je témou nášho príspevku.

Za kompozitá sme považovali všetky pomenovania vecí, vlastností a činností obsahujúce aspoň dva autosémantické kmene bez ohľadu na ich formu², ktoré pomenúvajú jeden denotát alebo obraz. Týmto spôsobom sa do nášho súboru dostali zloženiny typu *činohra*, *azúrovomodrý*, *banskobystrický*, *hrôzostrašný* atď. Cieľom príspevku bol lingvistický výskum kompozít, ich textových vlastností a ich vzájomných vzťahov, predovšetkým so zameraním na polytextiu.

Do skúmaného jazykového materiálu sme zaradili nasledujúce jazykové jednotky:

- a) čisté kompozitá – *zemegula*, *politicko-ekonomický*, *šéfredaktor*, *zemetrasenie*, *životopis*, *cukrovar*, *zákonodarca* atď.;
- b) abreviatúry – *RTVS*, *SIS*, *UNESCO*, *VOS*, *MOKYS*, *NBÚ* atď.;
- c) posesívne zloženiny – *čiernovláska*, *bielovlasý*, *tmavovláska* atď.;
- d) zrasty – *životunebezpečný*, *bohužiaľ*, *hrôzostrašný*, *mocipán*, *dôveryhodný* atď.;
- e) lexikalizované spojenia – *občianske právo*, *bravčové črevo*, *národný výbor*, *právny proces* atď.

Počas lingvistického výskumu sme si vytvorili a osvojili pojem „neanalyzované kompozitá“, ktorý označuje zloženiny nepatriace do našej výskumnej vzorky. Neanalyzovali sme číslovkové kompozitá, reduplikácie, zloženiny s cudzím základom a synsémantické spojenia. Tieto skupiny sme sa rozhodli vylúčiť z nášho jazykového súboru z nasledujúcich dôvodov:

² Ide o kompozitá, ktoré sa píše spolu alebo so spojovníkom, združené pomenovania, alebo lexikalizované spojenia s vysokou kohéziou, do ktorých sa nedajú vložiť ďalšie výrazy.

- a) číslovkové kompozitá – vysoká frekvencia by teoreticky mohla negatívne skresliť výsledky výskumu;
- b) reduplikácie – vysoká miera duplicity by pôsobila stereotypne;
- c) cudzojazyčné pomenovania – cudzí základ by komplikoval určovanie významov jednotlivých lexém;
- d) synsématické slovné spojenia – čiastočná alebo úplná neplnovýznamovosť by nespĺňala naše kritériá pri určovaní kompozít, za ktoré považujeme len spojenie dvoch plnovýznamových základov.

Vychádzame z predpokladu, že čím viac textových vlastností danej jednotky nájdeme a čím viac vzťahov medzi nimi preskúmame, tým presnejšie a komplexnejšie je možné danú jazykovú jednotku poznať. V príspevku sa zameriavame na vzťah medzi polytextiou a frekvenciou, ktorý zodpovedá nasledujúcim dvom hypotézam:

- a) čím je vyššia polytextia, tým je vyššia frekvencia zloženín;
- b) čím je vyššia frekvencia, tým je vyššia priemerná polytextia.

Vzťahy medzi polytextiou a frekvenciou sú vzájomne ovplyvňované, a teda každá z textových vlastností je ovplyvňovaná tou druhou, hoci nezáleží na tom, v akom jazykovom vzťahu sa nachádzajú.

a) Vzťah medzi polytextiou a frekvenciou môžeme merať dvomi smermi. Pri tomto predpoklade je polytextia nezávislou premennou a frekvencia závislou premennou. Ak vypočítame závislosť frekvencie na polytextii, môžeme sa riadiť obvyklou potenčnou krivkou danou ako:

$$Fr = a(Pt)^b, \text{ kde } a = \text{multiplikačná konštanta; } b = \text{potenčná konštanta.}$$

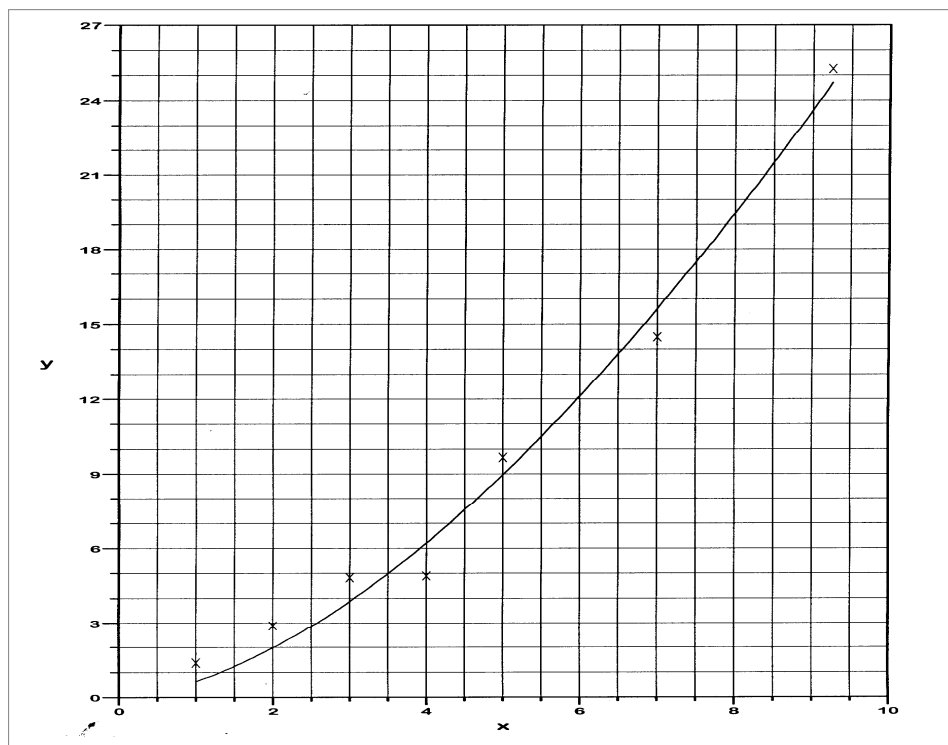
Uplatnením krivky dostaneme výsledok daný v Tabuľke 1. Krivka má formu $Fr = 0.6342(Pt)^{1.6463}$ a determinačný koeficient je $R^2 = 0.9895$, t. j. veľmi vysoký. Pri najvyšších hodnotách polytextie sme brali do úvahy priemery oboch premenných, to znamená, ak pri niektorých premenných bolo v tabuľke desatinné číslo, brali sme do úvahy ich priemery, čím sa zabezpečila reprezentatívnosť premenných.

Tabuľka 1: Závislosť frekvencie kompozít od polytextie

Polytextia (Pt)	Frekvencia (Fr)	Výsledná frekvencia
1.00	1.3738	0.6342
2.00	2.8846	1.9853
3.00	4.8214	3.8700
4.00	4.8889	6.2144
5.00	9.6667	8.9731
7.00	14.5000	15.6138
9.25	25.2500	24.7048

Výsledok, ktorý sme dosiahli pri skúmaní vzťahov medzi frekvenciou a polytextiou, je znázornený aj na nasledujúcom grafe.

Graf 1: Závislosť frekvencie kompozít od polytextie



V uvedenom grafe sa znázorňuje vzťah frekvencie k polytextii, kde x = polytextia a y = frekvencia. Ako ukazuje $R^2 = 0.9895$ a Graf 1, predpoklad o závislosti frekvencie od polytextie, t. j. čím je polytextia väčšia, tým je vyššia frekvencia, môžeme považovať za potvrdený.

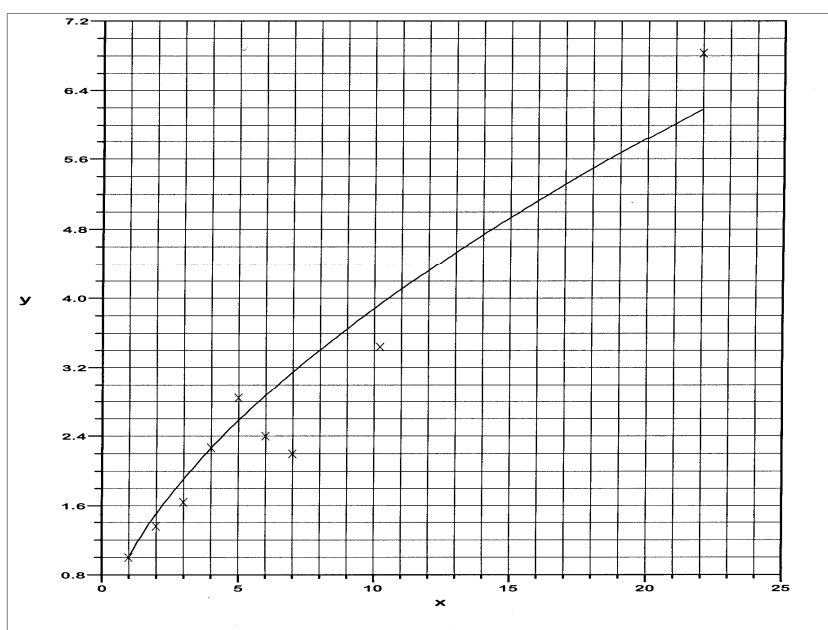
b) Vzťah medzi polytextiou a frekvenciou je možné merať aj opačným smerom – čím vyššia je frekvencia kompozít, tým vyššia je aj priemerná polytextia, hoci pri skúmaní vzájomných závislostí zvyčajne stačí merať závislosť vybraných vlastností len jedným smerom. Ak vypočítame naopak závislosť polytextie/závislá premenná od frekvencie/nezávislá premenná, tak dostaneme výsledok v Tabuľke 2.

Tabuľka 2: Závislosť polytextie kompozít od frekvencie

Frekvencia (Fr)	Polytextia (Pt)	Výsledná polytextia
1.00	1.0000	1.000000
2.00	1.3590	1.504767
3.00	1.6429	1.911095
4.00	2.2692	2.264325
5.00	2.8462	2.582683
6.00	2.4000	2.875753
7.00	2.2000	3.149340
10.22	3.4444	3.936517
22.00	6.8333	6.186036

Ako multiplikačnú konštantu sme dosadili $a = 1$ a dostali sme výsledok $Pt = (Fr)^{0,5895}$ s determináčným koeficientom $R^2 = 0.92$. Aj v tomto prípade sme pri vyšších hodnotách s desatinnými číslami pracovali s ich priemerom. Musíme poznamenať, že frekvencia nie je zárukou vyššej polytextie. Ak sa kompozitum veľmi často vyskytuje, môže sa vyskytovať len v jednom texte a nezávisle od polytextie. Frekvencia teda určuje polytextiu slabšie, ako je to naopak (pozri vyššie). Napriek tomu výsledok $Pt = (Fr)^{0,5895}$ ukazuje, že je to ešte veľmi vysoko signifikantné, čiže závislosť je zrejماً v oboch smeroch. Graficky tento vzťah znázorňuje Graf 2.

Graf 2: Závislosť polytextie kompozít od frekvencie



V uvedenom grafe sa znázorňuje závislosť polytextie od frekvencie kompozít, kde x = frekvencia a y = polytextia. Ako vidieť na grafe, krivka nie je taká pravidelná, na dvoch miestach sú výrazné odchýlky od krivky, ale napriek tomu signifikancia platí.

Záver: V matematickej lingvistike sa pri skúmaní textu uplatňujú metódy vychádzajúce z presnej definície merateľných jednotiek. Oproti kvalitatívnej lingvistike sa jazykové jednotky chápu z hľadiska ich fungovania v texte podľa vlastností, ktoré sa dajú formalizovať a dajú sa aplikovať na výskum viacerých jazykov. Preto aj pri našej definícii bolo určujúcim, že spojenie niekoľkých lexém opisuje len jeden jav alebo obraz. Vychádzali sme z racionálneho predpokladu o merateľnosti vlastností kompozít a ich vzťahu k modelovaniu a exaktným interpretáciám. Pri spracovaní danej problematiky sme využili Köhlerovu synergeticko-lingvistickú metodiku a softvérovú podporu podľa náhľadov a odporúčaní prof. G. Altmanna.

Predovšetkým sme v príspevku sústredili pozornosť na polytextiu zloženín a jej vzťah k frekvencii. Podstatným pre náš jazykový výskum sa stal aj zápis získaných exaktných údajov. Pomocou frekvencie sme určovali výskyt daného kompozita vo všetkých článkoch. Frekvenčné údaje sme získali excerpciou z 59 článkov periodika Plus 7 dní a dosiahnuté číselné hodnoty sme následne sčítali. Naopak pod polytextiou zloženín rozumieme výskyt jedného kompozita vo všetkých použitých textoch, resp. kontextoch. Môžeme tvrdiť, že polytextiu vnímame ako existenciu daného zloženého slova v článkoch, teda v koľkých publicistických textoch sa nachádza. Polytextia ako textová vlastnosť vystihuje frekvenciu určitej zloženiny a k nej príslušný počet článkov, resp. textov. Rozdiel medzi frekvenciou a polytextiou je nasledujúci:

- a) frekvencia je číselné vyjadrenie výskytu daného kompozita v textoch;
- b) polytextia je číselné vyjadrenie toho, v koľkých textoch sa dané kompozitum vyskytovalo.

Z výsledkov, ktoré sme dosiahli, sa ukazuje fakt, že frekvencia kompozít ovplyvňuje polytextiu a platí to aj naopak. Skúmanie vlastností ide zľava doprava, takže ak nejaká vlastnosť ovplyvňuje ďalšiu, naopak to platiť nemusí. V tomto prípade sme potvrdili postulované predpoklady, a tak sme prispeli do poznania fungovania vlastností kompozít v texte. Túto skutočnosť môžeme dokázať na konkrétnych príkladoch: čím je vyššia frekvencia, tým je vyššia polytextia, napr. kompozitum pravdepodobný má frekvenciu 12 a polytextiu má 11, na rozdiel od kompozita právomoc, ktoré má frekvenciu 1 a jeho polytextia je 1. Môžeme spomenúť aj kompozitum československý, ktoré má frekvenciu 4 a jeho polytextia je 3, na rozdiel od kompozita nocľah, ktoré má frekvenciu 1 a jeho polytextia je taktiež 1. Týmto spôsobom by sme mohli vymenovať viacero zloženín z výskumnej vzorky, ktorá obsahuje 625 kompozít.

Na záver môžeme konštatovať, že sa nám podarilo pozitívne overiť hypotézy a dokázať ich platnosť. Z výsledkov je zrejmé, že sa vybrané vlastnosti v texte nevyskytujú náhodne, ale vzájomne sa ovplyvňujú. Práve na toto overenie hypotéz sme využili štatistické metódy

a metódy modelovania kvantitatívnej lingvistiky, ktorá skúma jazyk pomocou štatistických metód so zameraním na frekvenciu jazykových jednotiek a javov. Výskum vzájomných vzťahov medzi textovými vlastnosťami kompozít je pomerne široká a otvorená téma, ktorá si zasluži ešte priestor.

Literatúra

BEST, Karl-Heinz: Linguistik in Kürze. Göttingen: Peust & Gutschmidt Verlag 2006. 154 s.

BOBEKOVÁ, Kristína: Polytextia kompozít. [diplomová práca] Trnava: Univerzita sv. Cyrila a Metoda 2012. 68 s.

KÖHLER, Reinhard: Zur linguistischen Synergetik. Struktur und Dynamik der Lexik. Bochum: Brockmeyer 1986. 201 s.

Zdroj

Plus 7 dní, 2011, roč. 21, č. 42 – 52.