

KVANTITATIVNÍ VÝZKUM VĚTY V RUŠTINĚ

(Na materiálu 4000 vět z novin)¹

Práce využívající statistických metod v jazykovědě mají těžiště ve zkoumání 1. frekvence slov, slovních druhů a tvarů. Umožňují např. lépe poznat uplatnění tzv. produktivních a neproduktivních tvarů, frekvenci slovních druhů apod. Takový výzkum má význam jak jako součást zevrubného popisu jazyka, tak i pro účely didaktické (výběr nejfrekventovanějších slov, volba postupu při vyučování gramatické — od sloves modálních, byť nepravidelných, avšak velmi frekventovaných apod.). Frekvenční seznamy slov a tvarů mohou také pomáhat k řešení některých otázek stylistických: statistická stylistika pomáhá jak při charakterizaci různých funkčních stylů, tak i při charakterizaci individuálních stylů autorských a mikrostylů;² 2. délky věty v jazyce. V tomto případě kvantitativní výzkum slouží dosud především jen k řešení otázek stylistických (viz práce Yulovy, Lesskisovy aj.).³

Zatímco závislost délky věty na okolnostech negramatických byla v poslední době předmětem mnoha studií, bylo dosud málo pozornosti věnováno vztahu délky věty k její gramatické stavbě. Jedním z velmi zajímavých perspektivních úkolů je prozkoumat, jakými prostředky dochází k rozšíření vět. Gramatická stavba dlouhé věty není totiž pouhým násobkem elementů krátké věty, neboť provedená výzkumná sonda ukazuje, že relativní frekvence různých slovních druhů a tvarů se s délkou věty velmi výrazně mění. Existují totiž a) prostředky specifické jen pro dlouhou větu (k nim patří především mezivětné spojky, participiální a přechodníkové vazby), b) prostředky typické pro dlouhou větu, které sice mohou být i ve větě krátké, ale s délkou věty jejich počet výrazně roste nejen v absolutním množství, ale i relativně vzhledem k ostatním elementům věty (sem patří především adnomi-nální genitiv, předložkové pády, adjektiva), c) prostředky, jejichž frekvence není závislá na délce věty, neboť jejich absolutní počet roste úměrně s délkou věty (sem patří především mezislovní spojky, akuzativní předmět, číslovky a predikativa), d) prostředky, jejichž relativní frek-

¹ Excerpovány byly tyto noviny: Правда, č. 195, r. 1965, s. 1—6; Правда č. 265, r. 1965, s. 1—6; Известия č. 242, r. 1965, s. 1—4; Известия č. 243, r. 1965, s. 1—3 (celkem 18¼ stran.

² Mám na mysli především základní práci v tomto směru; a to G. U. Yule, *On sentence length as a statistical characteristic of style in prose: with application to two cases of disputed authorship*, Biometrika 1939, sv. 30, s. 363.

³ G. A. Lesskis, *О размерах предложений в русской научной и художественной прозе 60-х годов XIX в.*, ВЯ 1962, č. 2, s. 78—95; *О зависимости между размером предложения и характером текста*, ВЯ 1963, č. 3, s. 92—112; *О зависимости между размером предложения и его структурой в разных видах текста*, ВЯ 1964, č. 3, 99—123.

vence klesá s délkou věty (sem patří např. zájmena, finitní tvary slovesné, částice a adverbia).

Problém délky věty není tedy výlučně záležitostí stylistickou, ale též závažným problémem syntaktickým. Při porovnávání jazyků můžeme zjišťovat i výrazné rozdíly ve volbě prostředků pro rozšíření věty.

*

Výběr 4000 vět z novin představuje úplný soubor po sobě následujících vět s tímto omezením: vynecháno bylo záhlaví listu, nadpisy, podpisy, místní a časové údaje, dále rubriky obsahující povětrnostní předpovědi, programy apod. Jazyk titulku představuje specifický styl novinářského projevu, maximálně koncentrovaného a reklamně zaměřeného. Z hlediska délky a frekvence slovních tříd představuje problematiku natolik odlišnou, že jeho vyloučení ze zkoumaného materiálu bylo nutné, aby se zabránilo zkreslení výsledných údajů o délce věty v novinách. Rovněž výrazně odlišný styl mají i uvedená stabilní sdělení o místě a dnu, autorovi, obsahu, programech apod.; u nich nejen slovní obsah, ale již samo umístění a grafická úprava plní sdělovací funkci a umožňuje maximální lakoničnost. Proto sdělení tohoto druhu představují z hlediska délky věty problematiku zcela samostatnou. Ale i po oddělení těchto typů jazykových projevů není jazyk novin jednotný, nýbrž představuje řadu funkčních mikrostylů. Rozlišují:

1. publicistika (autorizované zpravodajství, essaye, kritika)	1661 vět — 41,525 %
2. zpravodajství	1145 vět — 28,625 %
3. popularizace vědy a techniky	438 vět — 10,95 %
4. dopisy pracujících	267 vět — 6,675 %
5. překlady	221 vět — 5,525 %
6. úvodníky	160 vět — 4 %
7. projevy	108 vět — 2,7 %

Tyto mikrostyle se zřetelně odlišují jak průměrnou délkou věty (např. zatímco průměrný počet slov ve zkoumaném souboru vět je 15,115, v úvodníkové větě je to 18,287 slov), tak výskytem vět o určité délce.

Např. věty o 5 slovech (3,675 % vět daného souboru)	věty o 30 slovech (0,95 % vět daného souboru)
činí	0,374 %
7,116 % vět v dopisech pracujících	0,602 %
4,515 % vět v publicistice	1,659 %
2,882 % vět ve zpravodajství	—
2,262 % vět v překladech ⁴	—
2,078 % vět v popularizaci vědy a techniky	1,369 %
1,852 % vět v projevech	0,926 %
1,25 % vět v úvodních	0,625 %

⁴ Věty o 5 slovech se však vyskytly jen v překladu z lotyštiny. Naopak pro překlad z češtiny byly charakteristické věty dlouhé; také nejdelší věta z celého souboru vět byla v překladu z češtiny (měla 99 slov).

Jazyk novin představuje tedy jazykový projev dosti rozmanitý. Předpokládáme, že z hlediska cizojazyčného uživatele představuje jazyk novin nejběžnější druh jazykového projevu, s nímž přichází do styku.

* * *

Za pracovní vymezení slova jsem volila sled písmen mezi dvěma mezerami. Protože pro statistický výzkum je velmi důležitý postulát bezspornosti, viděla jsem při tomto rozboru slovo především jako jednotku grafickou.

Prakticky to znamená, že *написал бы* považuji za 2 slova, *он будет писать* za tři slova. Domnívám se, že není správné hodnotit jako jedno slovo např. složený předložkový výraz *в соответствии с, в связи с*, jak činí např. G. A. Lesskis⁵. Zkresluje se tím obraz o délce vět v určitém druhu textu, neboť také užívání podobných složených předložkových výrazů je jedním z prostředků prodlužování věty. Kde při vymezení hranice slova je rozpor mezi slovem jako jednotkou grafickou a významovou, vycházela jsem z grafické podoby. Pokud jde o slova spojená spojovací čárkou, vznikly tu jistě problémy. Spojovací čárka totiž nespojuje dvě slova v jedno, neboť i slova takto spojená mohou si zachovat ohebnost v obou částech. Teprve ztráta ohebnosti u prvního z nich umožňuje jim splýnout v jedno slovo. Proto jsem považovala slova spjatá spojovací čárkou za dvě slova v těch případech, kdy obě si zachovala ohebnost, např. *четверо его товарищей-шоферов* (3116), *отгружено угля-брака* (3152), *с заводами-изготовителями* (3063); za jedno slovo u slov neohybných nebo neskloňujících se, např. *на пол-очка* (2493), *с генерал-майором* (3409). Za dvě slova však považuji také např. *ракеты класса земля-воздух* (3834), *играет Отелло-негра*. Zde nesklonnost má jiné důvody. U slov cizích rozhodoval způsob psaní, i když užívání spojovací čárky bývá v ruštině u psaní cizích slov zcela nahodilé. (Srov. neodůvodněný rozdíl v psaní dvou slovanských měst jako *город Млада-Болеслав* (2174) — *Велико Тырново* (2217).)

Frekvence slovních druhů ve větě však sama o sobě málo říká o prostředcích uplatňujících se při rozšiřování vět. Např. frekvence substantiv neodhaluje způsob narůstání věty. To nám umožňuje teprve detailnější rozbor s přihlédnutím k funkčnímu využití. Pokud jde o substantiva, rozlišuji tedy frekvenci substantiv a) v nominativu, b) v bezpředložkovém akuzativu, c) v bezpředložkovém genitivu, d) v bezpředložkovém instrumentálu, e) v bezpředložkovém dativu, f) v předložkových pádech. Takové rozlišení dá vyniknout skutečnosti, že relativní počet genitivů v daném druhu textu stoupá s délkou věty nejvýrazněji, zejména v jednoduchých větách. Bližší průzkum pak ukazuje, že jde především o stoupající frekvenci adnominálního genitivu. Sr. *Вечером в Академическом театре оперы и балета имени Спендиарова мастера искусств Армении дали большой концерт в честь партийно-правительственной делегации ГДР*. (1403). Substantivum v kterékoliv větné pozici má totiž možnost být rozvíto řadou genitivů. Hromadění genitivů je pro daný druh textu mnohem příznačnější než hromadění přídavných jmen.

Podobně vidím nutnost rozlišit finitní tvary slovesné od infinitivu a participiálních tvarů. Pokud však jsou součástí složených tvarů slovesných, řadím

⁵ Zajímavý výklad této problematiky obsahuje stať V. G. Admonio, *Размер предложения и словосочетания как явление синтаксического строя*, ВЯ 1966, č. 4, 111—118.

je k finitnímu slovesu. To umožňuje pozorovat přibývání participií s délkou věty. Rozlišují také spojky mezislovní a mezivětné. Spojky, které stojí na začátku věty a plní funkci uvozovacích slov, řadím k částicím.

Za pracovní vymezení věty považují sled slov mezi dvěma tečkami, případně vykřičníky nebo otazníky, fungují-li jako koncové znaky věty (tj. musí-li následující obecné jméno začínat velkým písmenem). Po vykřičníku nebo otazníku se často píše malé písmeno, zvláště na konci přímé řeči, následuje-li pokračování věty uvozovací. Sr. — *A to еще спросит: „Что Вам пишут из дома?“ — говорит бывший кремлевский часовой.*

Věty člením co do složitosti na a) věty jednoduché, tj. takové, které mají 1 nebo více souřadně připojených přísudků, vztažených k témuž podmětu (bez ohledu na jejich délku), neopakovanému ani zájmenem; b) souvětí. Souvětí dále člením podle toho, kolik větných komponentů obsahují: věty o 2, 3, 4, 5, 6 a více větných komponentech.

Za souvětí nepovažují věty obsahující kromě hlavní věty pouze vsuvkové výrazy, nespjaté s větou spojovacími slovy, např. *да, ж сожалению, как видим, смешно сказать, как известно, как говорится, понимаете, каюсь, замечу, кажется* a pod. Avšak následuje-li po nich spojka, považují již takový celek za souvětí. Nepřítomnost spojky považují tedy za formální kritérium pro vydělení uvozovacích výrazů v ruštině.

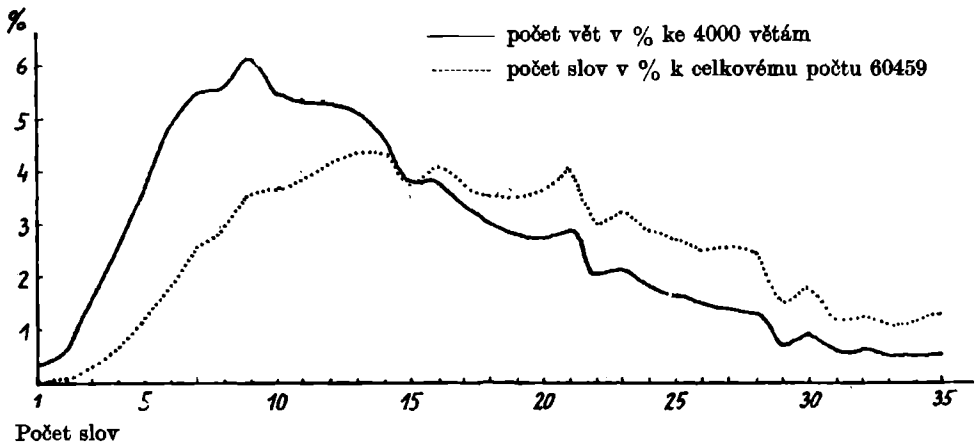
Kvantitativní údaje získané výzkumem

(Údaje týkající se mikrostylistického rozložení viz výše, str. 246).

Celkový počet slov v 4000 větách je	60 459
Průměrný počet slov ve větě je	15,115
Nejvíce je vět o 9 slovech	— 243 vět, tj. 6,1 %
Nejvíce slov připadá na věty o 13 slovech	— 2 665, tj. 4,412 %

Je důležité rozlišovat údaje o četnosti vět určité délky od údajů o množství slov připadajících na věty dané délky a údaje o průměrném počtu slov ve větě. Rozdíly mezi uvedenými ukazateli odrážejí pohyb křivky, znázorňující frekvenci vět podle jejich délky. Nejčastější

Graf č. 1



počet slov ve větě ukazuje maximální bod této křivky. Je-li nejčastější počet slov ve větě značně nižší než průměrná délka věty, obsahuje text více vět dlouhých. Rozdíl mezi nejčastějším počtem slov ve větě a délkou vět, které zahrnují nejvyšší podíl z celkového počtu slov, ukazuje, jak se pohybuje křivka frekvence vět se stoupajícím počtem slov. Přihlédnutí k počtu slov obsažených celkem v daném typu vět je mimořádně důležité proto, že velmi frekventované krátké věty mohou se podílet na celkovém textu méně významně než dlouhé málo frekventované. Např. v daném souboru věty o 1–9 slovech (tedy včetně nejfrekventovanějších vět!) neobsahují dohromady tolik slov jako co do počtu poměrně nečetné věty o 30 a více slovech. Rozložení vět co do četnosti slov charakterizuje graf č. 1 v % vzhledem k 4000 zkoumaných vět a k celkovému počtu 60 459 slov.

Vět ukončených otazníkem je 65, tj. celkem 1,625 % z celkového počtu. Mají celkem 715 slov, tedy průměrný počet slov v tázacích větách je 11. 40 tázacích vět ze 65 neobsahuje více než 10 slov.

Vět ukončených vykřičníkem je 75, tj. celkem 1,875 %. Mají dohromady 602 slov, tedy průměrný počet slov v těchto větách je 8,021. Avšak u 60 z celkového počtu nepřesahuje počet slov číslo 10.

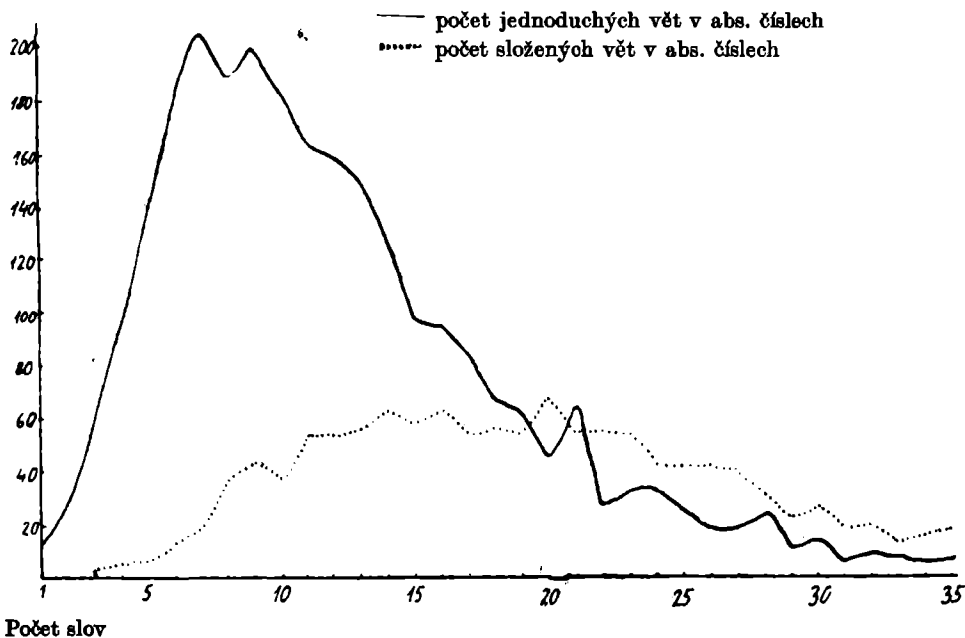
Věty podle složitosti:

Vět jednoduchých je	2687 — 67,175 %
Vět složených ze 2 větných komponentů	956 — 23,9 %
Vět složených ze 3 větných komponentů	261 — 6,525 %
Vět složených ze 4 větných komponentů	74 — 1,85 %
Vět složených z 5 větných komponentů	18 — 0,45 %
Vět složených z 6 a více větných komponentů	4 — 0,01 %

Výrazně tedy převládají věty jednoduché nad složenými. Poměr jednoduchých a složených vět znázorňuje graf č. 2.

Graf č. 2

Počet vět



Věty podle složitosti jsou zastoupeny co do délky takto:

Jednoduché věty podle jednotlivých tisícových souborů

	I.	II.	III.	IV.	Celkem:
Počet jednoduchých vět	653	701	647	686	2 687 vět
Počet slov	8307	8439	8371	8272	33 389 slov
Průměrný počet slov ve větě	12,73	12,04	12,9	12,06	12,42
Nejčastější počet slov	7	9	6	7	7
Nejvíce slov připadá na věty o	13	9	16	12	13 slovech

Složené věty podle jednotlivých tisícových souborů

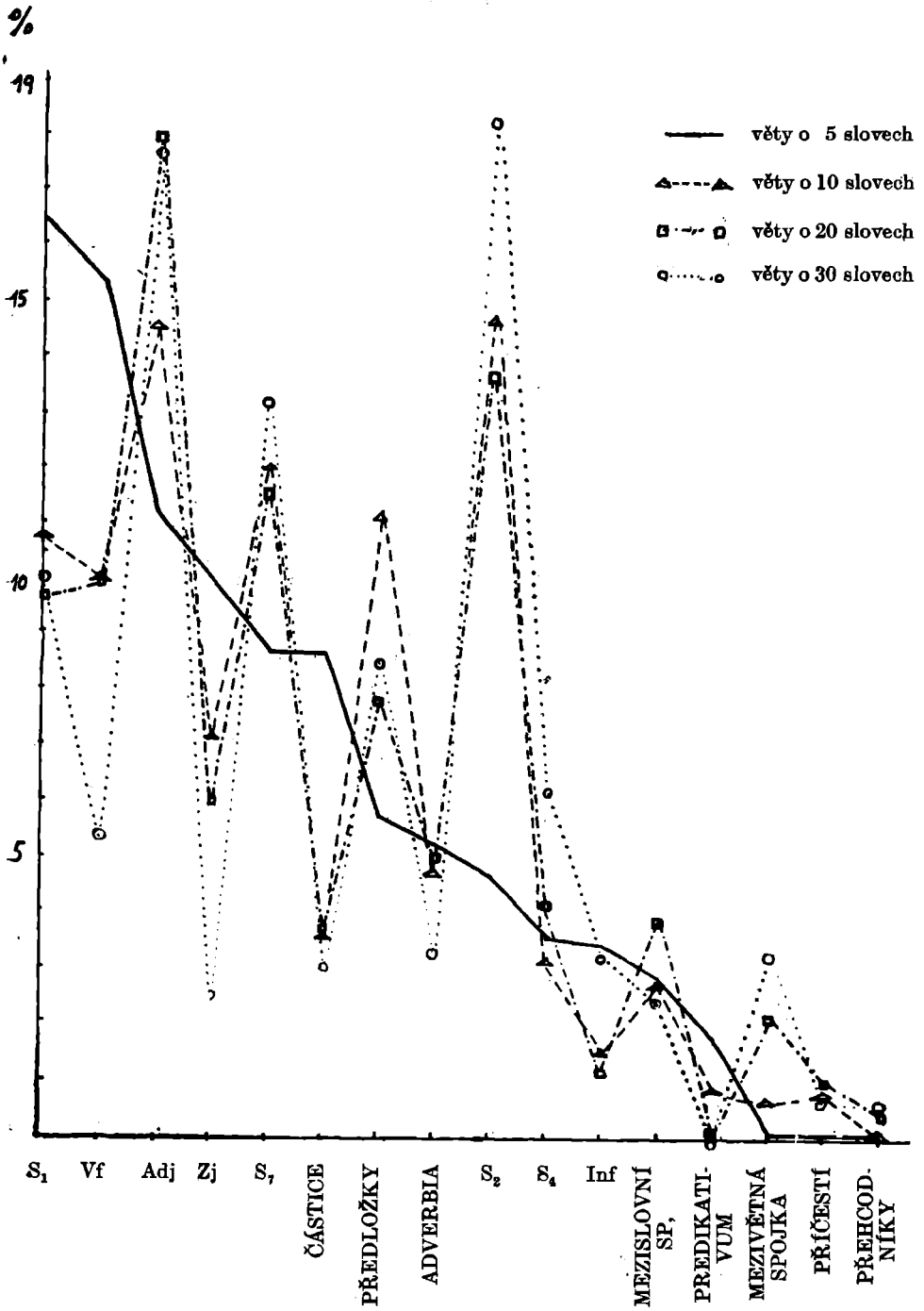
	I.	II.	III.	IV.	Celkem:
Počet slož. vět	347	299	353	314	1 313 vět
Počet slov	7319	5783	7581	6385	27 068 slov
Průměrný počet slov ve větě	21,075	19,34	21,5	20,3	20,6
Nejčastější počet slov	21	13	18	15—20	20
Nejvíce slov připadá na věty o	21	27	23	20	20 slovech

Značné diference mezi jednotlivými soubory jsou dány jednak tím, že v nich nejsou rovnoměrně zastoupeny všechny mikrostyle, jednak příliš malým počtem zkoumaných jevů (viz vzestup diferencí u složených vět).

Frekvenci slovních druhů a tvarů znázorňuje graf č. 3. Pro větší názornost a snazší zpracovatelnost pracovala jsem s takto omezeným počtem vět: vyšla jsem z nejméně frekventovanějšího rozsahu vět, tj. o 9 slovech. Protože jsem však potřebovala číslo dělitelné dvěma, vybrala jsem za základ věty o 10 slovech, dále věty o polovičním počtu slov — tj. o 5 slovech, věty o dvojnásobném počtu slov — tj. o 20 slovech, a konečně věty o trojnásobném počtu slov, tj. o 30 slovech. U takto vybraných vět jsem zkoumala poměrné zastoupení jednotlivých slovních druhů a tvarů podle výše uvedené klasifikace (srov. s. 247). Každý ze 4 tisícových souborů vět jsem zkoumala zvlášť a porovnávala výsledky. Zatím uvádím pouze závěry z průzkumu vět od 0—999, avšak mohu říci, že ani v dalších třech souborech se nejeví stav nijak podstatně odlišně.

Relativní výskyt jednotlivých slovních druhů a tvarů se mění s délkou věty. Výstavba věty dlouhé není pouze součtem prostředků, sloužících k výstavbě věty krátké. Věty dlouhé mají specifické prostředky pro rozšiřování. Jestliže uvážíme, že ve zkoumaném souboru vět tvoří jednoduché věty $\frac{2}{3}$ všech vět a že teprve mezi větami o 20 a více slovech převažují věty složené nad jednoduchými, pak tím více vynikne, že nejvýraznějším prostředkem k rozšiřování věty je adnominální genitiv. Jeho relativní výskyt neomezeně roste s délkou věty. Dalším prostředkem k narůstání věty je adjektivum. U vět velmi dlouhých však relativní počet adjektiv stagnuje, roste jen jejich absolutní počet. Kromě toho počet adjektiv je totiž v určitém vztahu k počtu substantiv ve větě. Počet adjektiv, jimiž může být rozvíjeno substantivum, je totiž gramaticky omezen, kdežto počet adnominálních genitivů ve větě gramaticky omezen není. Teprve na třetím místě mezi prostředky k rozšiřování věty jsou předložkové pády substantiv. Jejich relativní výskyt roste s délkou

Graf č. 3.



věty téměř tak rychle jako výskyt adjektiv, avšak na rozdíl od nich nejeví ani u velmi dlouhých vět tendenci k stagnaci, nýbrž dále roste. (Přiložený graf uvádí tyto předložkové pády s označením S_7 , avšak v tom jsou obsaženy i bezpředložkový dativ a instrumentál).

Frekvence předložek stoupá jen do omezené hranice; v daném případě relativní výskyt předložek u vět o 5—10 slovech se téměř zdvojnásobil, u vět o 20—30 slovech stoupal pouze absolutní počet předložek, avšak jejich relativní výskyt klesal. Je to dáno mimo jiné tím, že předložka se může vztahovat současně k několika bezprostředně následujícím substantivům. Zejména v psaném odborném projevu dochází tu k jisté schematizaci: předložka se vytýká před řadu substantiv podobně jako v matematice znaménko před závorky. Frekvence mezislovních spojek nijak výrazně nestoupá, ba u velmi dlouhých vět spíše mírně klesá. Je to způsobeno zásadami platnými pro užívání spojek, že jen mezi poslední dvě z řady souřadně připojených slov se klade spojka, zatím co v ostatních pozicích její funkci plní čárka. Frekvence adverbii s délkou věty relativně klesá. Nejsou tedy prostředkem k rozšiřování věty v daném souboru. Rovněž klesá, a to dosti výrazně, frekvence zájmen. Je to ovšem dáno stylem projevu. Podotýkám, že zájmena uvozující vztažné věty řadím k mezivětným spojkám.

V rozšiřování věty mají významný podíl participia, přechodníky a mezivětné spojky. Jsou to prostředky specifické pro dlouhou větu, a i když jejich relativní četnost je nepatrná, je vyvážena mimořádnou schopností být spojovacím článkem mezi hlavní predikací a relativně samostatnými predikacemi vedlejšími.

Uvedené závěry ukazují, že statistické zkoumání souvislostí ve frekvenci slovních druhů a tvarů s délkou věty je s to významně přispět k charakterizaci jak individuálních stylů autorských, tak i jazykových stylů (např. k charakterizaci stylu odborného) a konečně i k zpřesnění některých výsledků konfrontačního popisu jazyků. Uvedený výzkum, který je součástí širšího výzkumného programu oddělení východoslovanských jazyků při ÚJL pod vedením prof. A. V. Isačenka, představuje ovšem jen první sondu, jejímž účelem bylo mimo jiné zjistit, jak velký materiál je třeba k tomu, aby výsledné statistické údaje neobrážely nahodilé souvislosti.