

I. Kvantitativní lingvistika	7
1. Filozofické předpoklady	7
2. Pojetí kvantitativní lingvistiky	8
3. Předmět kvantitativní lingvistiky	10
II. Metodologická problematika	11
1. Jednotka souboru	12
1.1 Jednotka souboru v lexikální statistice	12
1.2 Jednotka souboru v gramatické statistice	15
1.2.1 Jednotka souboru v morfologické statistice	15
1.2.2 Jednotka souboru v syntaktické statistice	16
1.3 Jednotka souboru v sémantické statistice	17
1.4 Menší jednotky než slovo	18
1.5 Závěry	18
2. Výběr a rozsah materiálu (korpusu)	19
2.1 Výběr materiálu z hlediska kvalitativního	19
2.1.1 Kritéria jazyková	19
2.1.2 Kritéria psychologická	20
2.1.3 Kritéria sociologická	21
2.1.4 Kritéria tematická	22
2.1.5 Kritéria "sémiotická"	24
2.1.6 Jiná kritéria	26
2.2 Výběr a rozsah materiálu z hlediska kvantitativního	27
2.2.1 Způsob výběru materiálu	28
2.2.1.1 Systematický neboli mechanický výběr	28
2.2.1.2 Náhodný výběr	28
2.2.1.2.1 Náhodný výběr stránek	29
2.2.1.2.1.1 Náhodný výběr stránek v lexiku	30
2.2.1.2.1.2 Náhodný výběr stránek v gramatice	32
2.2.1.2.1.3 Využití náhodného výběru stránek	32
2.2.1.2.2 Náhodný výběr slov	33
2.2.1.3 Výběr souvislých částí textu	35
2.2.1.3.1 Způsob výběru souvislých částí textu	36
2.2.1.3.2 Výběr souvislých částí textu v lexikální statistice	36
2.2.1.3.3 Výběr souvislých částí textu v gramatické statistice	38
2.2.1.3.3.1 Výběr souvislých částí textu v morfologické statistice	38
2.2.1.3.3.2 Výběr souvislých částí textu v syntaktické statistice	40
2.2.1.3.3.3 Výběr souvislých částí textu v sémantické statistice	41
3. Statistické a jiné charakteristiky	42
3.1 Frekvence, rank, pořadí	42
3.1.1 Tzv. Zipfovy zákony	44
3.1.1.1 Tzv. první zákon Zipfův	44
3.1.1.2 Tzv. druhý zákon Zipfův	47
3.1.1.3 Tzv. třetí zákon Zipfův	48
3.1.1.4 Závěry	49
3.2 Průměr	49
3.3 Rozptyl a směrodatná odchylka	50
3.4 Rozdělení, rozložení, distribuce frekvencí (četností)	51
3.5 Koefficienty	53
3.5.1 Koefficient rozptýlení (disperze), ev. užití jednotky (slova)	54
3.6 Korelace, korelační koeficient	56
3.7 Míry teorie informace, entropie, redundance	58
3.8 Závěry	59
III. Základní oblasti kvantitativní lingvistiky	60
1. Lexikální statistika	60
1.1 Předmět lexikální statistiky	60
1.2 Metodologická problematika	61
1.2.1 Jednotka souboru a rozsah korpusu	61

1.2.2	Uspořádání slovní zásoby podle frekvenčního kritéria	63
1.2.2.1	Pásmo slov s frekvencí nejvyšší a vyšší	63
1.2.2.2	Pásmo slov s frekvencí střední	65
1.2.2.3	Pásmo slov s frekvencí nižší a nejnižší	66
1.2.2.4	Závěry	67
1.2.3	Tzv. bohatství slovníku	67
1.2.3.1	Formule P. Guirauda	67
1.2.3.2	Formule J. Mistriky	70
1.2.3.3	Formule M. Těšitelové	71
1.2.3.4	Závěry	72
1.3	Některé základní práce lexikální statistiky	73
1.3.1	do 60. let	73
1.3.2	od 60. let	73
1.3.3	Práce z lexikální statistiky týkající se jazyků slovanských	75
1.3.3.1	Čeština	75
1.3.3.2	Slovenština	76
1.3.3.3	Ruština a ukrajinština	77
1.3.3.4	Polština	78
1.3.3.5	Ostatní jazyky slovanské	79
1.4	Závěry	80
2.	Gramatická statistika	81
2.1	Předmět gramatické statistiky	81
2.2	Metodologická problematika	81
2.3	Rozdělení gramatické statistiky	82
2.3.1	Morfologická statistika	82
2.3.1.1	Předmět morfologické statistiky	82
2.3.1.2	Metodologická problematika	82
2.3.1.2.1	Jednotka souboru v morfologické statistice	82
2.3.1.2.2	Výběr materiálu v morfologické statistice	84
2.3.1.3	Některé základní práce v oblasti morfologické statistiky	88
2.3.1.3.1	Čeština	89
2.3.1.3.2	Slovenština	93
2.3.1.3.3	Ruština a jiné slovanské jazyky	93
2.3.1.4	Závěry	94
2.3.2	Syntaktická statistika	95
2.3.2.1	Předmět syntaktické statistiky	95
2.3.2.2	Metodologická problematika	95
2.3.2.2.1	Jednotka souboru v syntaktické statistice	95
2.3.2.2.2	Výběr materiálu v syntaktické statistice	97
2.3.2.3	Některé základní práce syntaktické statistiky	102
2.3.2.3.1	Čeština	103
2.3.2.3.2	Slovenština	105
2.3.2.3.3	Ruština a jiné slovanské jazyky	106
2.3.2.4	Závěry	107
3.	Sémantická statistika	107
3.1	Předmět sémantické statistiky	107
3.2	Metodologická problematika	108
3.2.1	Jednotka souboru v sémantické statistice	108
3.2.2	Výběr materiálu a statistických metod v sémantické statistice ..	110
3.3	Některé základní práce v oblasti sémantické statistiky	112
3.4	Závěry	115
IV.	Ostatní oblasti kvantitativní lingvistiky	116
1.	Fonologická statistika	116
1.1	Předmět fonologické statistiky	116
1.2	Metodologická problematika	116
1.2.1	Jednotka souboru ve fonologické statistice	116
1.2.2	Výběr a rozsah materiálu	117
1.3	Některé základní práce fonologické statistiky	118
1.3.1	Čeština	119
1.3.2	Slovenština	120
1.3.3	Ruština a ostatní slovanské jazyky	121

1.4 Závěry	120
2. Grafematická statistika	122
2.1 Předmět grafematické statistiky	122
2.2 Metodologická problematika	122
2.2.1 Jednotka souboru v grafematické statistice	123
2.2.2 Výběr a rozsah materiálu	123
2.3 Některé základní práce grafematické statistiky	124
2.3.1 Čeština	124
2.3.2 Ostatní jazyky slovanské i jiné	125
2.4 Závěry	126
3. Stylistická statistika	126
3.1 Předmět stylistické statistiky	126
3.2 Metodologická problematika	127
3.2.1 Jednotka souboru v stylistické statistice	127
3.2.2 Výběr a rozsah materiálu	128
3.3 Některé základní práce stylistické statistiky	129
3.3.1 Vlastní stylistické práce	129
3.3.1.1 Čeština	130
3.3.1.2 Slovenština	131
3.3.1.3 Ruština a jiné slovanské jazyky	132
3.4 Závěry	133
4. Typologická statistika	134
4.1 Předmět typologické statistiky	134
4.2 Metodologická problematika	134
4.2.1 Jednotka souboru v typologické statistice	134
4.2.2 Výběr a rozsah materiálu	134
4.3 Některé základní práce typologické statistiky	135
4.4 Závěry	136
5. Jiné oblasti kvantitativní lingvistiky	136
5.1 Vývoj jazyka/jazyků	137
5.1.1 Předmět kvantifikace	137
5.1.2 Metodologická problematika	137
5.1.3 Některé základní práce týkající se vývoje jazyka	138
5.2 Další oblasti kvantitativní lingvistiky	139
V. Kvantitativní lingvistika a počítače	140
VI. Aplikace výsledků kvantitativní lingvistiky v praxi	144
VII. Závěry	148
Rejstřík nejdůležitějších pojmů a termínů kvantitativní lingvistiky	149
Seznam literatury	161