LEXICOGRAPHY

BY

LEW R. MICKLESEN

1.10 The Scientific Word Counts

In this section four different word counts are presented and discussed. The first of these is a high-frequency word count in which all semantic units are listed which occurred 10 or more times in the 31,403-word corpus. The second word count is a rearrangement of the first word count: the high-frequency semantic units are listed according to form class. The third word count is concerned with all semantic units of multiple form class which occurred in the high-frequency list: a tabulation is made of the number of times each of these semantic units was used for each particular form class.

The fourth word count is concerned with the English alternatives for each Russian semantic unit in the high-frequency list: the numbers of times each alternative was chosen in the corpus is listed.

1.11 The High-Frequency Word Count

The high-frequency word count presented in Word Count I lists all semantic units which occurred 10 or more times in the corpus of 111 Russian selections. The 341 members of this high-frequency list account for 11,752 of the 31,403 occurrences of semantic units in the text.

A detailed discussion of the high-frequency list, together with a comparison of the list with the general-language list compiled by Josselson is presented in Chapter 1 of R. E. Wall's thesis, which may be found elsewhere in this report.

И	1197	МОЖНО	49	Т.Е.	29
В	1076	ЯВЛЯЕТСЯ	48	БЫЛО	28
НА	470	ОНА	47	ОДНОЙ	28
С	316	КОТОРЫЕ	44	ЧТОБЫ	28
ПРИ	305	ЗА	42	ЭТОМ	28
НЕ	248	ПОЭТОМУ	41	ВСЕГДА	27
ДЛЯ	220	МОГУТ	40	ИМЕЮТ	27
К	197	ТАК КАК	40	КОТОРОЙ	27
ПО	192	НАПРИМЕР	39	РАЗЛИЧНЫХ	27
ОТ	188	ПОД	39	БОЛЬНЫХ	26
ИЛИ	185	ЧЕМ	38	ВЕЛИЧИНЫ	26
ИЗ	182	ЧЕРЕЗ	38	ДВИЖЕНИЯ	26
ЧТО	154	ВРЕМЯ	37	ПОЧВЫ	26
А	143	ОБЫЧНО	37	ПРИ ЭТОМ	26
ИХ	118	ЯВЛЯЮТСЯ	37	СЛУЧАЯ	26
КАК	108	НИХ	36	ВОЗДУХА	25
У	105	ПОСЛЕ	36	ВСЛЕДСТВИЕ	25
ДО	95	ЕЩЕ	35	ОСОБЕННО	25
О	84	БУДЕТ	33	ВЕСА	24
ЕГО	79	ВОДЫ	33	ВСЕХ	24
БОЛЕЕ	66	ВРЕМЕНИ	32	КОЛИЧЕСТВО	24
ТО	66	ИМЕЕТ	32	ЛИШЬ	24
НО	65	СЛЕДУЕТ	32	СВЯЗИ	24
МОЖЕТ	60	НАИБОЛЕЕ	31	ТОМ	24
ТОЛЬКО	58	НЕОБХОДИМО	31	ЭТОЙ	24
ЕЕ	56	ЭТИХ	31	ДАЖЕ	23
ЕСЛИ	56	ВСЕЙ	30	ЗНАЧИТЕЛЬНО	23
МЕЖДУ	56	СИСТЕМЫ	30	ОН	23
ТАКЖЕ	56	ТАК	30	БЕЗ	22
ЭТО	55	ЭТИ	30	ВОЛН	22
ЭТОГО	53	ОЧЕНЬ	29	ГДЕ	22
ВСЕ	51	СЛУЧАЕ	29	ДЕЙСТВИЯ	22
МЫ	51	ТЕЛА	29	ДРУГИХ	22

ЖЕ	22	ЛУЧЕИ	15	НИЖЕ	12
КОТОРЫЙ	22	ОПЕРАЦИИ	15	ОБРАЗОМ	12
ПУТЕМ	22	ОШИБКИ	15	ОСНОВНЫХ	12
ТАКИМ ОБРАЗОМ	22	ПРИ ПОМОЩИ	15	ОТДЕЛЬНЫХ	12
ФИЛЬТРАЦИЙ	22	ПРОЦЕСС	15	ПОЛКА	12
ЧАСТИ	22	СИЛЫ	15	ПРИМЕРНО	12
ЗНАЧЕНИЕ	21	СОДЕРЖАНИЕ	15	РИС.	12
КИСЛОТА	21	ТОКА	15	СИСТЕМЕ	12
ЛИНИИ	21	УСЛОВИЙ	15	СКОРОСТЬ	12
РАБОТЫ	21	ФУНКЦИИ	15	ТЕМПЕРАТУРЫ	12
РАСТЕНИЙ	21	ЧАСТИЦ	15	УСЛОВИЯ	12
СЛЕДОВАТЕЛЬНО	21	ЭТОТ	15	ФОРМ	12
ФУНКЦИЙ	21	БОЛЬШЕ	14	БУДУТ	11
БАТАЛЬОНА	20	БОЛЬШИХ	14	ВЕЛИЧИНА	11
ВЕСЬМА	20	БЫЛИ	14	ВИД	11
ВИДЕ	20	ВЕС	14	ВЛИЯНИЕ	11
НЕСКОЛЬКИХ	20	ГЛАВНЫМ ОБРАЗОМ	14	ВПОЛНЕ	11
НЕСКОЛЬКО	20	ГРУППЫ	14	ДАВЛЕНИЯ	11
СТВОЛА	20	ДВА	14	ДАННЫЕ	11
ВОЗМОЖНОСТЬ	19	КЛЕТОК	14	ДРУГИЕ	11
ДОЛЖНЫ	19	ЛЕНТЫ	14	ЖИВОТНЫХ	11
И Т. Д.	19	НАПРЯЖЕНИЯ	14	ЗНАЧЕНИЯ	11
МЕНЕЕ	19	НИМ	14	ИЗВЕСТНО	11
ОБЛАСТИ	19	ОСНОВНОЙ	14	ИМЕЮТСЯ	11
ОДНАКО	19	ОТНОСИТЕЛЬНО	14	ИССЛЕДОВАНИЯ	11
ПРИЧЕМ	19	ПИТАНИЯ	14	КОЛИЧЕСТВА	11
СТОРОНЫ	19	ПЛОСКОСТИ	14	КОРАБЛЕЙ	11
ТЕМ	19	ПОВЕРХНОСТЬ	14	КОТОРОМ	11
ТОЧКИ	19	ПРИМЕНЕНИЕ	14	МЕНЬШЕ	11
ЭНЕРГИИ	19	РАБОТА	14	МЕСТО	11
ВМЕСТЕ	18	СРЕДИ	14	НЕКОТОРЫЕ	11
ДАЕТ	18	УЖЕ	14	ОДИН	11
ЗАВИСИТ	18	УСЛОВИЯХ	14	ОСИ	11
МАСЛА	18	ФОРМЫ	14	ПЕРВОГО	11
НЕКОТОРЫХ	18	БЫ	13	ПОЛЕ	11
НЕОБХОДИМО	18	БЫЛ	13	ПРЕИМУЩЕСТВЕННО	11
ОДНОВРЕМЕННО	18	ВИДЕТЬ	13	ПУТИ	11
ПОЧТИ	18	ВОПРОС	13	РАССМАТРИВАТЬ	11
СКОРОСТИ	18	ДВИЖЕНИЕ	13	САМО	11
СТРОЯ	18	ДОЛЖЕН	13	СИСТЕМУ	11
БОЛЬШОЙ	17	ДОЛЖНА	13	СЛУЧАЕВ	11
ВЫШЕ	17	ЗАДАЧА	13	СТЕПЕНИ	11
ГАЗЫ	17	ИЗМЕРЕНИЙ	13	ТАКОЙ	11
ЗАТЕМ	17	ИМЕТЬ	13	УЧАСТКА	11
ПРОИСХОДИТЬ	17	КОТОРАЯ	13	ХАРАКТЕРИЗОВАТЬ	11
ПРОТИВНИКА	17	КОТОРОГО	13	ЧАЩЕ	11
СОБОЙ	17	КРОМЕ	13	ЧЕРТ	11
СРАВИТЕЛЬНО	17	КРЫЛА	13	ЭЛЕКТРОДА	11
ТЕЧЕНИЕ	17	НАСТОЯЩЕЕ	13	АЛЮМИНИЯ	10
ЗДЕСЬ	16	ОНА	13	АТАКИ	10
ИНОГДА	16	ПРИРАЩЕНИЕ	13	БАКТЕРИИ	10
НАЧАЛЬНИКА	16	РЕЗКО	13	БАКТЕРИЙ	10
СМ.	16	РЕЗУЛЬТАТЕ	13	БОЛЬШОЕ	10
СТРУКТУРЫ	16	СВЕТА	13	БЫТЬ	10
ТОГО	16	СОСТАВ	13	ВНУТРИ	10
ФИГ.	16	ТЕХ	13	ВНУТРЬ	10
ЧАСТО	16	УГОЛ	13	ВЫЗЫВАТЬ	10
ХАРАКТЕР	16	ЧАСТИЦЫ	13	ВЫСАДКИ	10
ШТАБА	16	В РЕЗУЛЬТАТЕ ЧЕГО	12	ВЫСОТЫ	10
Я	16	ВЕЩЕСТВА	12	ГАЗОВ	10
БЫЛА	15	ВИДЫ	12	ГРУППА	10
БЫСТРО	15	ВЫСОКОЙ	12	ДАЛЕЕ	10
ВСЕГО	15	ДЕТАЛЕЙ	12	ДРУГИМ	10
ДАВЛЕНИЕ	15	КОЛЕБАНИЯ	12	ЗАКЛЮЧАЕТСЯ	10
ДВИГАТЕЛЕЙ	15	КОТОРЫЙ	12	ИЗУЧЕНИЯ	10
ДВУХ	15	ЛИБО	12	И Т. П.	10
ЖИДКОСТЬ	15	МАТЕРИАЛА	12	КИСЛОТЫ	10
ИЗМЕНЕНИЯ	15	МНОГИХ	12	КОСТИ	10
КЛЕТКИ	15	МОЩНОСТИ	12	МАШИНЫ	10
ЛЕГКО	15	НЕПОСРЕДСТВЕННО	12	МЕРЕ	10

МЕСТА	10	ПРИВОДА	10	СРЕДНЕЙ	10
МЕТОД	10	ПРОТЕЗА	10	СРЕДЫ	10
МЕТОДЫ	10	РАЗЛИЧНЫЕ	10	ТАКИХ	10
НАД	10	РАССТОЯНИЕ	10	ТИПА	10
ОДНОГО	10	РЕЖИМА	10	ТКАНЕЙ	10
ОНО	10	СИЛЬНО	10	ТРЕБУЕТСЯ	10
ОТСУТСТВИЕ	10	СИСТЕМА	10	ХОРОШО	10
ПЕРИОД	10	СОЕДИНЕНИЯ	10	ЦВЕТ	10
ПЕЧИ	10	СОСТАВА	10	ЭНДОСПЕРМА	10
ПОТОКА	10	СПОСОБЫ	10	ЭТИМ	10
ПРЕДСТАВЛЯЮТ	10	СРЕДЕ	10	ЕМУ	10
ПРЕЖДЕ ВСЕГО	10				

1.12 The High-Frequency Word Count According to Form Classes

The second word count is concerned with a tabulation of the semantic units of the high-frequency list according to their form class. For the purposes of this study, 28 form classes were defined according to Table I. Semantic units of the high-frequency list belonged to only 21 of the 28 form classes; i.e., no possessive adjectives, intensive adjectives, participles (long form), gerunds, interjections, prefixes, or punctuation marks are tabulated in the high-frequency list.

A total of 146 substantives occurred in the high-frequency list. The most frequent, время, occurred 37 times, while 34 of the 146 occurred 10 times each. It is interesting to note that the most frequent semantic unit, и, occurred 1197 times. Hence, while the substantive list has by far the largest membership, the most frequent substantive occurs less than 1/30 as often as the most frequent word. This is demonstrated in another way by observing that the total number of occurrences of the 146 substantives in the high-frequency list is 2174. The substantives therefore constitute $\frac{146}{341}$ (100) = 43% of the number of members in this list, but only account for $\frac{2174}{11,752}$ (100) = 18.5% of the occurrences in the corpus of members of this list.

A total of 21 prepositions occurred in the high-frequency count, but these 21 prepositions accounted for 3672 of the 11,752 occurrences. The prepositions thus consisted of $\frac{21}{341}$ (100) = 6.2% of the total number of words, but accounted for $\frac{3672}{11,752}$ (100) = 31.2% of the occurrences.

A complete tabulation of the number of members in each form class, together with the occurrences of all the members of each form class, is presented in Table I.

WORD COUNT NO. II

Distributional Classes

1. Substantives
2. Pro-substantives
3. Descriptive adjectives
4. Interrogative-indefinite adjectives
5. Possessive adjectives
6. Demonstrative adjectives
7. Intensive adjectives
8. Limiting adjectives
9. Ordinals
10. Numerals
11. Finite verbs
12. Principal infinitives
13. Auxiliary infinitives
14. Participles (long form)
15. Adjectival complements (short form of adjectives and participles)
16. Impersonal expressions (verbs and verbal adverbs)
17. Gerunds
18. Prepositions
19. Adverbs
20. Coordinating conjunctions
21. Subordinating conjunctions
22. Particles
23. Interjections
24. Prefixes
25. Punctuation
26. Multiple distribution class
27. Idiomatic sequences
28. Parenthetic word or clause

Distributional Class No. 1--Substantives

37	время	21	кислота	16	начальника	15	частиц
33	воды	21	линии	16	структуры	14	вес
32	времени	21	работы	16	фиг.	14	группы
30	системы	21	растений	16	характер	14	клеток
29	случае	21	функции	16	штаба	14	ленты
29	тела	20	батальона	15	давление	14	напряжения
26	величины	20	виде	15	двигателей	14	питания
26	движения	20	ствола	15	жидкость	14	плоскости
26	почвы	19	возможность	15	изменения	14	поверхность
26	случая	19	области	15	клетки	14	применение
25	воздуха	19	стороны	15	лучей	14	работа
24	веса	19	точки	15	операции	14	условиях
24	количество	19	энергии	15	ошибки	14	формы
24	связи	18	масла	15	процесс	13	вопрос
22	волн	18	скорости	15	силы	13	движение
22	действия	18	строя	15	содержание	13	задача
22	фильтрации	17	газы	15	тока	13	измерений
22	части	17	противника	15	условий	13	приращение
21	значение	17	течение	15	функций	13	результате
						13	света

WORD COUNT NO. II (continued)

13	состав	11	вид	10	алюминия	10	период
13	угол	11	влияние	10	атаки	10	печи
13	частицы	11	давления	10	бактерии	10	потока
12	вещества	11	значения	10	бактерий	10	привода
12	виды	11	исследования	10	высадки	10	протеза
12	деталей	11	количества	10	высоты	10	расстояние
12	колебания	11	кораблей	10	газов	10	режима
12	материала	11	место	10	группа	10	система
12	мощности	11	оси	10	изучения	10	соединения
12	образом	11	поле	10	кислоты	10	состава
12	полка	11	пути	10	кости	10	способы
12	рис.	11	систему	10	машины	10	среде
12	системе	11	случаев	10	мере	10	среды
12	скорость	11	степени	10	места	10	типа
12	температуры	11	участка	10	метод	10	тканей
12	условия	11	черт	10	методы	10	цвет
12	форм	11	электрода	10	отсутствие	10	эндосперма
11	величина						

Distributional Class No. 2--Pro-Substantives

51	мы	23	он	16	я	13	она
47	они	17	собой	14	ним	10	оно
36	них						

Distributional Class No. 3--Descriptive Adjectives

27	различных	14	основной	12	основных	10	средней
17	большой	13	настоящее	12	отдельных	10	различные
14	больших	12	высокой	10	большое		

Distributional Class No. 4--Interrogative-Indefinite Adjectives

		18	некоторых	11	некоторые		

Distributional Class No. 6--Demonstrative Adjectives

55	это	28	этом	15	этот	10	таких
53	этого	24	этой	13	тех	10	этим
31	этих	16	того	11	такой	10	эту
30	эти						

Distributional Class No. 8--Limiting Adjectives

30	всей	24	всех	11	один	10	одного
28	одной	15	всего				

Distributional Class No. 9--Ordinals

22	других	11	другие	11	первого	10	другим

Distributional Class No. 10--Numerals

20	нескольких	15	двух	14	два		

Distributional Class No. 11--Finite Verbs (including Auxiliary)

60	может	32	следует	18	необходимо	11	будут
48	является	28	было	17	происходит	11	имеются
40	могут	27	имеют	15	была	10	заключается
37	являются	18	дает	14	были	10	представляют
33	будет	18	зависит	13	был	10	требуется
32	имеет						

Distributional Class No. 12--Principal Infinitives

13	видеть	13	иметь	11	рассматривать	11	характеризовать

Distributional Class No. 13--Auxiliary Infinitives

		10	быть	10	вызывать		

Distributional Class No. 15--Adjectival Complement

19	должны	13	должен	13	должна		

Distributional Class No. 16--Impersonal Expressions

49	можно						

Distributional Class No. 18--Prepositions

1076	в	192	по	84	о	22	без
470	на	188	от	56	между	22	путем
316	с	182	из	42	за	14	среди
305	при	105	у	38	через	13	кроме
220	для	95	до	25	вследствие	10	над
197	к						

Distributional Class No. 19--Adverbs

66	более	29	очень	19	менее	16	здесь
56	также	27	всегда	18	вместе	16	иногда
41	поэтому	25	особенно	18	одновременно	11	вполне
35	еще	22	где	18	почти	11	преимущественно
31	наиболее	20	весьма	17	затем	10	далее

Distributional Classes No. 20 and 21--Coordinating and Subordinating Conjunctions

185	или	65	но	28	чтобы	12	либо
143	а	56	если	21	следовательно		

Distributional Class No. 22--Particles

248	не	22	же	13	бы		

Distributional Class No. 26--Multiple Form

1197	и	31	необходимо	19	тем	12	которым
154	что	30	так	17	выше	12	непосредственно
118	их	29	то есть	17	сравнительно	12	ниже
108	как	27	которой	16	см.	12	примерно
79	его	26	больных	16	часто	11	данные
66	то	24	том	15	быстро	11	известно
58	только	23	даже	15	легко	11	котором
56	ее	23	лишь	14	больше	11	меньше
51	все	22	значительно	14	относительно	11	чаще
44	которых	22	который	14	уже	10	внутри
41	которые	22	путем	13	которого	10	внутрь
40	после	20	несколько	13	крыла	10	высоко
39	под	19	однако	13	резко	10	сильно
38	чем	19	причем	12	которая	10	хорошо
37	обычно						

Distributional Class No. 27--Idiomatic Sequences

40	так как	22	таким образом	14	главным образом	10	прежде всего
26	при этом	15	при помощи	12	в результате чего		

Distributional Class No. 28--Parenthetic Word or Clause

39	например	19	и т. д.	10	и т. п.

1.13 The Frequency of Occurrence of the Constituent Form Classes of Multiple-Form-Class Words

Word Count III is a tabulation of the multiple form class. The multiple form class has 55 members in the high-frequency list. As may be seen in Table II, the multiple form class is large in both membership and occurrences: members of the multiple form class account for 16% of the semantic units and 24% of the occurrences in the high frequency list.

Word Count III shows that 51 of the 55 members of the multiple form class have two form class possibilities, while the remaining 4 members have three possibilities. It may be said, therefore, that semantic units of the multiple form class, generally, may have structural assignments according to either of two different form classes. Usually one form class assignment is much more likely than the other: the list shows that 33 of the 55 semantic units had one particular form class assignment in more than 90% of the occurrences.

	Total frequency	Substantives	Pro-substantives	Descriptive adjectives	Interr.-Indef. adjs.	Possessive adjs.	Demonstr. adjs.	Limiting adjs.	Numerals	Adj. complements	Prepositions	Adverbs	Conjunctions	Particles	Parenthetical exprs.	Finite & aux. verbs	Participles	Impersonal expressions
и	1197	--											1180	17				
что	154	--	2										152					
их	118	--	47			71												
как	108	--										38	70					
его	79	--	35			44												
то	66	--					14						52					
только	58	--										47	3	8				
ее	56	--	23			33												
все	51	--	1					50										
которых	44	--	44		0													
которые	41	--	41		0													
под	39	-- 0									39							
чем	38	--	4										34					
обычно	37	--								1		36						
после	36	--									36	0						
необходимо	31	--								4						27		
так	30	--										3	23	4				
т.е.	29	--																
который	27	--	27		0													
больших	26	-- 25		1														
лишь	24	--										24	0					
том	24	-- 0					24											
даже	23	--											3	20				

WORD COUNT NO. III (continued)

	Total frequency	Substantives	Pro-substantives	Descriptive adjs.	Interr.-Indef. adjs.	Possessive adjs.	Demonstr. adjs.	Limiting adjs.	Numerals	Adj. complements	Prepositions	Adverbs	Conjunctions	Particles	Parenthetical exprs.	Finite & aux. verbs	Participles	Impersonal exprs.
значительно	23	--								1		22						
который	22	--	22		0													
несколько	20	--							8			12						
однако	19	--											18	1				
причем	19	--										16	3					
тем	19	-- 0					19											
выше	17	--		3								14						
сравительно	17	--								0		17						
см.	16	-- 9														7		
часто	16	--								0		16						
быстро	15	--								0		15						
легко	15	--								4		11						
больше	14	--		6								8						
относительно	14	--								1	6	7						
уже	14	--		0								14						
которого	13	--	13		0													
крыла	13	-- 13														0		
резко	13	--								0		13						
которая	13	--	13		0													
которым	12	--	12		0													
непосредственно	12	--								0		12						
ниже	12	--		2								10						
примерно	12	--								0		12						
данные	11	-- 11														0		
известно	11	--								0								11
меньше	11	--		3								8						
чаще	11	-- 1		0								10						
внутри	10	--									7	3						
внутрь	10	--									3	7						
сильно	10	--								0		10						
хорошо	10	--								0		10						

1.14 The Frequency of Individual Alternatives According to Form Classes

Word Count No. 4 presents all source-language semantic units with a frequency of occurrence of 10 or more and with two or more target-language alternatives according to their constituent form classes and tabulates the frequency of occurrence for each alternative. The presentation therefore proceeds in the following hierarchy: 1) form class, 2) total frequency, 3) frequency of individual alternatives. The semantic units marked with an asterisk are multiple-form class units. They are listed under all form classes concerned. The frequency count for the alternatives of such equivalents will be listed only once under the appropriate form class. In place of a frequency count for alternatives that are not relevant there will be an indication of which form class will supply their counts. No particular significance is attached to this list. When the target-language equivalents were being assigned to the Russian semantic units, a conscious effort was made to list first those

alternatives intuitively felt to occur most frequently. The reader is invited to check the accuracy of this assignment.

SUBSTANTIVES

	Word	Equivalent	Total		1	Frequency by equivalents 2	3	4	5	6	7	8	9	10
	время	time/tense$_{62}$	37	--	35	2								
	времени	time/tense$_{62}$	32	--	30	2								
	случае	case/chance/occurrence	29	--	29	0	0							
	величины	magnitude/quantity	26	--	13	13								
	движения	motion/movement	26	--	2	24								
*	больных	sick/patients	26	--	adj	25								
	почвы	soil/ground	26	--	20	6								
	связи	connection/communication/bond$_3$/coupling$_{63}$/brace$_{94}$/liaison$_{97}$	24	--	15	5	4	0	0	0				
*	том	volume/that	24	--	0	dem adj								
	действия	act/action/effect/operation	22	--	0	13	1	8						
*	путем	by means/way/passage/journey/track$_{93}$	22	--	prep	0	0	0	0					
	части	part/department/fate/unit$_{97}$	22	--	22	0	0	0						
	значение	significance/value	21	--	12	9								
	ствола	trunk/barrel/shaft$_{91}$	20	--	3	17	0							
	области	area/oblast	19	--	19	0								
	стороны	side/direction/aspect	19	--	10	4	5							
*	тем	that/those/themes	19	--	dem adj	dem adj	0							
	точки	point/period/sharpening	19	--	19	0	0							
	масла	butter/oil	18	--	3	15								
	строя	line/formation/system/operation/tuning/constructing/drawing up	18	--	0	17	0	1	0	0	0			
	противника	enemy/opponent	17	--	17	0								
	течение	current/course	17	--	8	9								
*	см.	see/cm.	16	--	verb	9								
	штаба	staff/headquarters	16	--	9	7								
	клетки	cage/square/cell/mesh$_{91}$	15	--	0	2	13	0						
	процесс	process/disease$_{53}$/suit$_{68}$	15	--	15	0	0							
	силы	force/strength/effort	15	--	11	3	1							
	содержание	contents/maintenance/salary	15	--	15	0	0							
	тока	current/mating-place	15	--	15	0								
	клеток	cages/squares/cells$_4$/meshes$_{91}$	14	--	0	0	14	0						
	ленты	ribbon/tape/band/sliver-lap$_{98}$	14	--	7	0	0	7						
	напряжения	tension/effort/stress/voltage	14	--	2	3	0	9						
	питания	feeding/nourishment	14	--	12	2								
	плоскости	plane/flatness/surface	14	--	11	0	3							
	формы	form/uniform	14	--	14	0								
	движение	motion/movement	13	--	5	8								
*	крыла	wing/blade/fender/wall$_{73}$/covered	13	--	13	0	0	0	vb					
	света	light/world/society	13	--	13	0	0							
	состав	composition/staff/amount/compound$_{55}$	13	--	13	0	0	0						
	угол	corner/angle	13	--	0	13								
	виды	views/shapes/species$_4$/aspects$_{62}$	12	--	0	9	3	0						
	деталей	details/articles	12	--	1	11								
	колебания	oscillation/hesitation	12	--	12	0								
	образом	form/way/image	12	--	0	12								
	полка	regiment/shelf/weeding	12	--	12	0	0							
	форм	forms/uniforms	12	--	12	0								
	величина	magnitude/quantity	11	--	8	3								
	вид	view/shape/species$_4$/aspect$_{62}$	11	--	0	9	0	2						
	влияние	influence/effect	11	--	2	9								
*	данные	data/given	11	--	11	part								
	значения	significance/value$_1$	11	--	1	10								
	кораблей	ships/naves$_{94}$	11	--	11	0								
	оси	axis/axle	11	--	10	1								
	поле	field/floor/sex/brim/margin/flap	11	--	11	0	0	0	0	0				
	пути	way/passage/journey/track$_{93}$	11	--	4	1	5	1						
	случаев	cases/chances/occurrences	11	--	11	0	0							
	степени	degree/power$_1$	11	--	11	0								
	участка	area/lot	11	--	11	0								

SUBSTANTIVES

	Word	Equivalent	Total		Frequency by equivalents 1	2	3	4	5	6	7	8	9	10
*	чаще	thicket/more frequent/more thick/ more rapid/more frequently/more thickly/ more rapidly	11	--	1	adj	adj	adj	adv	adv	adv			
	высадки	landing/getting off/setting out/ transplants$_{46}$	10	--	10	0	0	0						
	высоты	height/altitude/pitch	10	--	6	4	0							
	кости	bone/die	10	--	10	0								
	машины	machine/machinery	10	--	10	0								
	печи	stove/furnace	10	--	0	10								
	потока	torrent/stream/flow	10	--	0	0	10							
	привода	bringing/drive	10	--	0	10								
	режима	condition/process/regime$_{65}$/ regimen$_{54}$/method	10	--	7	3	0	0	3					
	соединения	combination/connection/compound$_3$/ coupling$_{81}$/unit$_{97}$	10	--	0	1	4	5	0					
	состава	composition/staff/amount/compound$_{55}$	10	--	8	2	0	0						
	среде	medium/environment/Wednesday	10	--	8	2	0							
	среды	medium/environment/Wednesday	10	--	9	1	0							
	тканей	cloths/tissues	10	--	0	10								
	цвет	color/flower	10	--	10	0								

PRO-SUBSTANTIVES

*	что	that/what/which	154	--	conj	2	conj							
*	их	their/theirs/them	118	--	pos adj	pos adj	47							
*	его	his/its/him/it	79	--	pos adj	pos adj	0	35						
*	ее	her/it/hers/its	56	--	0	23	pos adj	pos adj						
*	чем	what/which/something/than/the	38	--	2	2	0	conj	conj					
	он	he/it	23	--	3	20								
	собой	self/selves	17	--	11	6								
	ним	him/it/them	14	--	1	5	8							
	она	she/it	13	--	2	11								

DESCRIPTIVE ADJECTIVES

*	больных	sick/patients	26	--	1	subst								
*	выше	higher/taller/before	17	--	3 adv	0	adv							
*	больше	more/bigger	14	--	adv	6								
	больших	big/bigger	14	--	14	0								
*	уже	already/more narrow/more narrowly	14	--	adv	0	adv							
	высокой	high/tall	12	--	12	0								
*	ниже	lower/shorter/below	12	--	2	0	adv							
	отдельных	separate/individual	12	--	6	6								
*	меньше	smaller/less	11	--	3	adv								
*	чаще	thicket/more frequent/more thick/ more rapid/more frequently/more thickly/more rapidly	11	--subst	0	0	0	0	adv	adv	adv			
	средней	middle/average/mean$_1$/neutral$_3$/ neuter$_{62}$	10	--	6	2	2	0	0					

INTERROGATIVE-INDEFINITE ADJECTIVES

	некоторых	certain/some	18	--	3	15								
	некоторые	certain/some	11	--	4	7								

DEMONSTRATIVE ADJECTIVES

*	то	that/then/or	66	--	14	conj	conj							
*	том	volume/that	24	--subst		24								
*	тем	that/those/themes	19	--	19	0	subst							

LIMITING ADJECTIVES

	одной	one/alone	28	--	28	0								

LIMITING ADJECTIVES

Word	Equivalent	Total	1	Frequency by equivalents 2	3	4	5	6	7	8	9	10
один	one/alone	11 --	11	0								
одного	one/alone	10 --	10	0								

NUMERALS

* несколько	several/somewhat	20 --	8	adv								

FINITE VERBS

является	is/appears	48 --	48	0								
являются	are/appear	37 --	37	0								
будет	will/will be	33 --	13	20								
следует	it is necessary/follows/is due	32 --	29	3	0							
* необходимо	it is necessary/is necessary/ necessary	31 --	27	3	adj							
дает	gives/allows	18 --	17	1								
происходит	takes place/comes	17 --	17	0								
* см.	see/cm.	16 --	7	subst								
* крыла	wing/blade/fender$_{91}$/wall$_{73}$/ covered	13 --	sub	sub	sub	sub	0					
будут	will/will be	11 --	4	7								
заключается	is concluded/ is contained/ is confined	10 --	10	0	0							
требуется	it is required/is required	10 --	4	6								

PRINCIPAL INFINITIVES

рассматривать	consider/observe	11 --	9	2								

ADJECTIVAL COMPLEMENTS

* обычно	is usual/usual/usually	37 --	0	1	adv							
* необходимо	it is necessary/is necessary/ necessary	31 --	aux verb	3	1							
* значительно	considerable/significant/considerably/significantly	23 --	1	0	adv	adv						
должны	must/should/owe	19 --	13	6	0							
* сравнительно	comparative/comparatively	17 --	0	adv								
* часто	frequent/frequently/thick/thickly/ rapid/rapidly	16 --	0	adv	0	adv	0	adv				
* легко	light/easy/lightly/easily/it is easy/it easy	15 --	0	2	adv	adv	2	0				
* относительно	concerning/relative/relatively	14 --	0	3 prep	adv							
должен	must/should/owe	13 --	13	0	0							
должна	must/should/owe	13 --	9	4	0							
* резко	sharp/sharply	13 --	0	adv								
* непосредственно	direct/directly	12 --	0	adv								
* примерно	exemplary/approximate/approximately/exemplarily	12 --	0	0	adv	adv						
* сильно	strong/strongly	10 --	0	adv								
* хорошо	good/it is good/ it good/ well/ all right	10 --	0	0	0	adv	0					

IMPERSONAL EXPRESSION

можно	is possible/it is possible/ it possible	49 --	2	41	6							

PREPOSITIONS

в/во	in/to/at/on/of/like/by	1076 --	910	15	33	4	18	0	2			
на	on/in/at/to/for/by/with	470 --	243	84	46	43	43	16	0			
с/со	with/from/about/on	316 --	278	34	0	4						

PREPOSITIONS

	Word	Equivalent	Total		1	Frequency by equivalents 2	3	4	5	6	7	8	9	10
	при	at/with/before/in time of/during/in/when/on	305	--	39	112	0	53	65	24	6	6		
	к/ко	to/toward/for/with	197	--	160	24	11	2						
	по	on/by/along/for/in/after/to	192	--	24	65	52	12	34	4	1			
	от	from/of/for/on/Ø	188	--	169	10	0	8	1					
	из	out/from/of	182	--	23	107	52							
	у	at/by/with/from	105	--	8	3	93	1						
	до	to/up to/before	95	--	41	50	4							
	о/об,обо	about/against/with/of	84	--	76	0	0	8						
	между	between/among	56	--	49	7								
	за	behind/beyond/for/after/in/at/under/by/with/because of	42	--	7	4	20	2	5	2	1	0	1	0
*	под	under/near/like/hearth-bottom	39	--	38	1	0	subst						
	через	over/through/within/every/every other	38	--	1	32	5	0	0					
*	путем	by means/way/passage/journey/track	22	--	22	subst	subst	subst	subst					
*	относительно	concerning/relative/relatively	14	--	0	6 adj com	adv							
	среди	among/in middle of	14	--	13	1								
	над	over/above	10	--	7	3								

ADVERBS

*	как	how/as/but	108	--	3	35 conj	conj							
*	обычно	is usual/usual/usually	37	--	adj com	adj com	36							
	еще	still/yet/more/also	35	--	19	7	1	8						
	особенно	especially/peculiarly	25	--	23	2								
*	лишь	only/as soon as	24	--	24	conj								
*	значительно	considerable/significant/considerably/significantly	23	--	adj com	adj com	20	2						
*	несколько	several/somewhat	20	--	num	12								
*	выше	higher/taller/before	17	--	3 adj	adj	11							
*	сравнительно	comparative/comparatively	17	--	adj com	17								
*	часто	frequent/frequently/thick/thickly/rapid/rapidly	16	--	adj com	16	adj com	0	adj com	0				
*	легко	light/easy/lightly/easily/it is easy/it easy	15	--	adj com	adj com	0	11	adj com	adj com				
*	больше	more/bigger	14	--	8	adj								
*	относительно	concerning/relative/relatively	14	--prep		adj com prep	7							
*	уже	already/more narrow/more narrowly	14	--	14	adj	0							
*	резко	sharp/sharply	13	--	adj com	13								
*	непосредственно	direct/directly	12	--	adj com	12								
*	ниже	lower/shorter/below	12	--	adj	adj	10							
*	примерно	exemplary/approximate/approximately/exemplarily	12	--	adj com	adj com	12	0						
*	меньше	smaller/less	11	--	adj	8								
*	чаще	thicket/more frequent/more thick/more rapid/ more frequently/ more thickly/more rapidly	11	--	sub	adj	adj	adj	10	0	0			
*	сильно	strong/strongly	10	--	adj com	10								
*	хорошо	good/it is good/it good/well/all right	10	--	adj com	adj com	adj com	10	adj com					

CONJUNCTIONS

*	и	and/even/too/also/indeed/both	1197	--1109		particle	16	56	particle	particle				
	а	but/and	143	--	28	115								
*	как	how/as/but	108	--	adv	70 adv	0							
*	то	that/then/or	66	--	dem adj	49	3							
*	чем	what/which/than/the/something	38	--	pro-sub	pro-sub	26	8	0					
*	лишь	only/as soon as	24	--	adv	0								

WORD COUNT NO. IV (continued)

PARTICLES

Word	Equivalent	Total	1	2	3	4	5	6	7	8	9	10
				Frequency by equivalents								
и	and/even/too/also/indeed/both	1197 --	conj	8	conj	conj	4	5				
же	however/then/also/very	22 --	12	7	3	2						
бы	would have/would/Ø	13 --	1	12	0							
IDIOMATIC SEQUENCES												
при этом	besides/at this/with this/before this/in time of this	26 --	11	0	10	0	5					
главным образом	chiefly/by main image/with main image/as main image/main image	14 --	14	0	0	0	0					

1.15 Total Occurrences of Target-Language Equivalents Containing Alternatives With Subscript Numbers

Since the subscript numbers associated with target-language alternatives relate the alternatives concerned to the Synoptic Chart of Fields of Science and Technology[1], this count, in effect, tabulates the number of occurrences of technical terms in the 111 text passages. This count is of some interest, at least, in corroborating the statement that "technical terms are rare islands in an ocean of general language."[2]

The data are presented first according to whether all of the alternatives of equivalents have subscript numbers or not, then according to form class.

A. All alternatives possess subscript numbers:
   1. Substantives 6
   2. Descriptive adjectives . . . 5
   TOTAL 11

B. Not all alternatives have subscript numbers:
   1. Substantives 1593
   2. Descriptive adjectives . . 227
   3. Verbs 167
   4. Multiple form 51
   5. Adverbs 5
   TOTAL . . . . . 2043 plus 11 = GRAND TOTAL: 2054.

There are 2054 occurrences of equivalents with alternatives bearing subscript numbers in the simulated machine translations of the 111 text passages. The total occurrences of all equivalents in the 111 text passages is 30,403. This means that the terms originally selected by the project investigators as technical terms accounted for only 6.8% of the total number of occurrences.

NUMBER OF SEMANTIC UNITS

Number of article	Number of words	Number of article	Number of words	Number of article	Number of words	Number of article	Number of words
1 --	434	16 --	269	31 --	486	46 --	144
2 --	205	17 --	291	32 --	291	47 --	194
3 --	293	18 --	313	33 --	387	48 --	287
4 --	364	19 --	223	34 --	376	49 --	364
5 --	260	20 --	380	35 --	359	50 --	363
6 --	215	21 --	346	36 --	314	51 --	256
7 --	236	22 --	344	37 --	212	52 --	238
8 --	194	23 --	316	38 --	289	53 --	202
9 --	402	24 --	243	39 --	285	54 --	417
10 --	350	25 --	364	40 --	294	55 --	133
11 --	271	26 --	296	41 --	299	56 --	243
12 --	340	27 --	163	42 --	297	57 --	301
13 --	317	28 --	193	43 --	324	58 --	356
14 --	217	29 --	292	44 --	340	59 --	281
15 --	389	30 --	444	45 --	197	60 --	213

[1] See Procedural Report, Section 12.5 and Automatic Pinpointing of Intended Non-Grammatical Meaning, Section 8.1, in Univ. of Wash. report, "Linguistic and Engineering Studies in Automatic Language Translation of Scientific Russian Into English," 1958.

[2] Erwin Reifler, Report on the First Conference in Mechanical Translation published in "Mechanical Translation," Vol. I, No. 2, August 1954, p. 25, column 2.

NUMBER OF SEMANTIC UNITS (continued)

Number of article	Number of words	Number of article	Number of words	Number of article	Number of words	Number of article	Number of words
61 --	214	74 --	212	87 --	210	100 --	314
62 --	406	75 --	95	88 --	352	101 --	314
63 --	283	76 --	188	89 --	213	102 --	268
64 --	311	77 --	211	90 --	142	103 --	379
65 --	330	78 --	316	91 --	205	104 --	137
66 --	198	79 --	234	92 --	161	105 --	369
67 --	204	80 --	291	93 --	505	106 --	386
68 --	243	81 --	191	94 --	169	107 --	531
69 --	208	82 --	160	95 --	256	108 --	460
70 --	261	83 --	268	96 --	326	109 --	309
71 --	169	84 --	213	97 --	200	110 --	398
72 --	255	85 --	112	98 --	349	111 --	293
73 --	129	86 --	186	99 --	475		31,403

1.20 An Experiment in the Automatic Selection[3] or Rejection of Technical Terms

One of the products of the early stages of machine translation work at the University of Washington was the so-called Synoptic Chart for Fields of Science and Technology reproduced on page 84 of this section. This chart was discussed in the first report[4] of the University of Washington Project. Whenever a given Russian semantic unit seemed to belong to a specific subfield, i.e., to one and only one of the rectangles on the chart, or to a specific field, i.e., to an entire vertical column in the chart, it was given an appropriate number from the chart. This number was to appear as a subscript numeral of the English alternative concerned and was to guide the reader of the output in his selection or rejection of technical terms on the basis of an awareness of the field of science represented by the subject matter of the text. The assignment of various alternatives to fields and subfields of science and technology and the classification itself of fields and subfields could not be checked until the simulated machine translations were produced. Once these translations became available the checking was fairly simple but extremely time-consuming. The first and critical step in the procedure was the careful perusal of every so-called text passage in the original corpus of the University of Washington MT Project and the subsequent assignment of the contents to one or more of the fields and subfields of science and technology. Next, the simulated machine translation for every text passage was scanned for equivalents containing alternatives bearing subscript numbers. If the subscript number coincided exactly with the number or one of the numbers assigned the text passage or if the one-digit subscript number corresponded to the first digit of the number or one of the numbers of the text passage, then the alternative associated with the subscript number was treated as if it had been automatically selected as the correct one. All alternatives bearing subscript numbers that did not correspond to numbers assigned to text passages in either of the ways described above were considered to be automatically rejected as if they were incorrect.

Obviously, the selection and rejection of alternatives on this basis had to be evaluated; therefore each selection or rejection was simultaneously evaluated against the context for correctness or incorrectness. Thus, each alternative with a subscript was tested against two sets of oppositions: selection-rejection and correct-incorrect. All these data were recorded in great detail for easy reference and subsequent evaluation. A convenient summary of the data listing the number of the text passage, the field of science number or numbers assigned to the text passage, and the numbers of correct and incorrect selections and rejections is presented on page 85. The total number of decisions is 2944 (correct decisions = 2588 incorrect decisions = 356). This means that 88% of the decisions were correct. The procedure was not entirely automatic since the initial classification of a text passage according to its field of science was made by a human being. It will be some time before the stage is reached when this classification can be effected by a machine as it scans a technical article.

After the data had been properly recorded, the incorrect decisions were subjected to analysis in the hope that they might provide information leading to the improvement in the form and application of the synoptic chart of fields of science and technology. The dichotomy of the incorrect decisions into incorrect rejections and incorrect selections proved to be very significant in the analysis; so the results of the analysis will be discussed first in terms of incorrect rejections, later in terms of incorrect selections.

The analysis of the incorrect rejections revealed that in the vast majority of cases no adjustment was necessary in the assignment of a text passage to a field of science. The analysis also revealed that the assignment of subscript numbers to alternatives, insofar as it allowed only one number per alternative, was correct. Nevertheless, the assignment was inadequate in that it was too specific. The incorrect rejections resulting from too high specificity may be classified into two groups on the basis of whether the adjustment necessary to correct the fault involved merely a reduction of the degree of specificity or the complete removal of specificity.

---

[3] This paper has been presented at the National Symposium in Machine Translation, February 2-5, 1960 at UCLA. For the collection of this material I am indebted to Mr. Friedrich Lackmann.

[4] See Footnote No. 1.

	1	2	3	4	5	6	7	8	9
0	Math.	Physics	Chemistry	Biology	Medicine	Social Sciences	Integrated Sciences	Applied Sciences	Technology
1	Algebra	Classical and Fluid Mechanics	Physical Chemistry	Botany	Structure	Anthropology	Astronomy	Mechanics & Mechanical Structures	Machinery Mechanism Tools
2	Geometry	Statistical Mechanics and Thermo	Inorganic Chemistry	Zoology	Function	Linguistics	Geophysics	Thermo & Heat Engines	Production and Mfg. Methods
3	Analysis	Electricity and Magnetism	Analytical Chemistry	Microbiology	Diagnosis	Philosophy	Geology	Electrical Engineering	Transportation
4	Statistics	Optics Spectra	Organic Chemistry	Biophysics	Therapy	Sociology	Geography	Aeronautical Eng. Acoustics	Structures Architecture
5	Numerical Analysis	Quantum Mechanics	Biochemistry	Psychology	Pharmacy	Pol. Sci. Diplomacy	Meteorology	Nuclear Engineering	Mining Metals Ceramics
6	Relativity	Solid State	Photo-Chemistry	Agriculture and Forestry	Public Health Sanitation	Social Planning	Oceanography	Control Engineering	Marine and Naval
7		Nucleonics	Electro-Chemistry	Animal Husbandry	Psychiatry	Economics Theoretical & Applied		Optics Photography	Military Science Tactics
8		Metrology	Chemical Engineering	Fisheries	Veterinary	Law		Materials	Textiles Paper
9									

Note: A Single digit subscript denotes a whole field, i.e., 2 for physics. A two-digit subscript denotes a subfield, i.e., 23 for electricity and magnetism as a subfield of physics.

SYNOPTIC CHART FOR SCIENCE AND TECHNOLOGY

SUMMARY OF CORRECT AND
INCORRECT SELECTIONS AND
REJECTIONS ACCORDING TO TEXT PASSAGES

Text Passage	Field of Science Number Assigned Text Passage	Correctly Rejected	Correctly Selected	Incorrectly Rejected	Incorrectly Selected
1	28	23		4	
2	13	18			2
3	13	20	11		
4	15	13	9		3
5	21/22	15			
6	21	18		1	
7	11	8	3		4
8	12	16			14
9	43	20	3	1	1
10	51	103	18	5	1
11	51	19	1	14	
12	54	17	7	1	
13	43	14	15		
14	54	18	4		
15	53	61	13	2	
16	54	14		1	
17	54	18		10	
18	54	5	1	1	
19	52	19	3		
20	57	21			
21	41	35	6	2	
22	41	27	6		4
23	41	37	8		3
24	41	23			
25	41	27		1	
26	42	10	1		5
27	42	10	1		
28	58	16	2	3	
29	42	21			
30	42	41	5		1
31	42	38	7	9	
32	52/ 42/58	20	10		3
33	42	35	14		5
34	42	43	2	5	
35	42	31	5	2	
36	46	7			4
37	46	6			
38	46	14			
39	46	16	4	1	1
40	73	9		1	
41	73	23			
42	73	35		12	2
43	73	12	3	2	1
44	73	14			
45	74	10			
46	73	19			
47	72	4		1	
48	74	15			
49	75	30			
50	75	20		4	
51	38/88	14			
52	35	23			
53	33/73	4	1		1
54	34	26	4	1	2
55	31/72	9			1
56	34	17			
57	88	12		1	
58	35	12		1	
59	23/72	10			
60	28	6			1
61	22/72	10		3	1
62	84	11		2	
63	27	17			
64	84/24	38	3		
65	25	22			
66	25	36		3	
67	23	13			
68	71	14	2	3	
69	46	11	1		2
70	91	18	2		19
71	91	7			
72	91	25		7	3
73	91	22	7	1	5
74	91	27	5	1	3
75	91	1	2		2
76	91	22	2	3	
77	72	3		2	
78	91/86	22	1	10	6
79	95/88	11			4
80	95/88	20	5	2	11
81	92/88	25		2	4
82	95	9	3		
83	92	13		1	
84	92	10		12	2
85	92	6			2
86	92	14	1		3
87	94	13	1		
88	97	19	16		1
89	92	8			
90	92	29			
91	92	19			
92	84	20		1	
93	84	49	21	2	
94	84	22		5	
95	83	58	8	2	2
96	83	27	18	2	2
97	83	21	4		1
98	83	17	2	1	1
99	62	15	10		2
100	62	17			3
101	65	41		5	
102	65	20			
103	97	10	2		2
104	97	19	10		9
105	96	21		11	1
106	96	59	2		4
107	96	58	1	9	1
108	97	23	1	1	16
109	97	21	1		2
110	97	30	2		3
111	84	6		3	
TOTALS		2,288	300	180	176

There were two primary areas on the Synoptic Chart where a reduction in degree of specificity of subscript assignments seemed to be particularly effective in eliminating incorrect rejections. Vertical columns 4 and 5 (biology and medicine, respectively) constitute the first area. More than one-fourth of all the incorrect rejections are apparently due to the fact that the subfields of biology and medicine are not as distinct from each other as those in column 6 (social sciences) and column 7 (integrated sciences). A large common vocabulary is shared by most or all branches of medicine and biology, and it seems inadvisable to restrict most words to one specific field or subfield. The following is a partial list of Russian semantic units with alternatives bearing too specific subscript numbers from columns 4 and 5:

зрительный	=	optic$_{5}$	слезный	=	lachrymal$_{5}$
лоскут	=	graft$_{54}$	слой	=	lamella$_{4}$
оболочка	=	membrane$_{51}$	срез	=	section$_{4}$
пинцет	=	forceps$_{5}$	узел	=	ganglion$_{51}$
проток	=	duct$_{51}$	чувствительный	=	sensory$_{4}$
связка	=	ligament$_{51}$			

Vertical columns 8 and 9 (applied sciences and technology) constitute the second area. About one-ninth of all the incorrect rejections seem due to the fact that these two technical areas are not always distinct. Their vocabularies frequently overlap. The following list includes a partial representation of Russian semantic units with alternatives bearing too specific subscript numbers from columns 8 and 9:

ввод	=	lead-in$_{83}$	пояс	=	flange$_{9}$
вкладыша	=	bushing$_{9}$	простой	=	demurrage$_{93}$
муфта	=	clutch$_{81}$/coupling$_{81}$/-socket$_{81}$	расчетный устройство	=	rated$_{83}$ working-principle$_{91}$

Two types of solutions suggest themselves for remedying these incorrect rejections. The most obvious method of reducing specificity is to increase the area on the Synoptic Chart to which given alternatives apply. This can be done by adding the number for an entire vertical column or part of another column, or it can be accomplished by adding the number for part or the rest of the same column. The number of subscript numbers employed certainly constitutes a factor; it would not seem advisable to use more than two such numbers. In the case of the two areas discussed above, columns 4 - 5 and 8 - 9, the decisions do not seem too difficult. Here it seems feasible to give words common to both 4 and 5 or 8 and 9 double subscript numbers, e.g., клетка = cell$_{4}$, ткань = tissue$_{4}$ could have number 5 as well as number 4, and вскрытие = dissection$_{5}$ and покров = integument$_{51}$ could have number 4 in addition to number 5 and 51. In instances where alternatives with 2-digit subscripts are not shared by 4 - 5 or 8 - 9 but have wide currency within a single column they should be reevaluated to determine whether or not they might be provided with single-digit (columnar) subscripts. A number of alternatives in these technical areas are undoubtedly specific enough to be permitted either single subscript numbers or two-digit subscript numbers. For example, угнетение = depression$_{5}$ and опухоль = tumor$_{53}$ can certainly be considered medical terms rather than biological; and most of the terms with subscript numbers 96 (marine and naval) and 97 (military science and tactics) are certainly distinct from the terms associated with other subfields in columns 8 and 9.

The other method of attacking the problem of too high a degree of specificity of present subscript numbers is to reevaluate the synoptic chart itself. Again the columns 4 - 5 and 8 - 9 provide an excellent case in point. It is very possible that a careful reconsideration of these columns of the chart might counsel that the fields of medicine and biology, on one hand, and the fields of applied sciences and technology, on the other hand, could subsequently be classified into appropriate subfields which, in turn, would reflect more accurately the distribution of technical terms. Parts of columns may also require consideration and reclassification. Two text passages discussing naval science contained a number of incorrect rejections because the alternatives concerned bore subscript number 97 (military science and tactics). This situation suggests that one solution may lie in the reevaluation of these two subfields with the possible creation of another subfield where the two overlap.

In all there were 180 incorrect rejections. If the above suggestions for reduction of specificity were to prove successful, 51% (92 out of 180) of the incorrect rejections would be avoided.

As might be expected, incorrect rejections remedied apparently by the complete removal of specificity were not confined to any particular areas on the Synoptic Chart. Incorrect rejections in this category are all alternatives denoting concepts used extensively in science and technology -- more or less general scientific terminology. The following is a list of Russian words with alternatives which had been erroneously limited to one field or subfield of science.

уравнение	=	equation$_{1}$	осаждать	=	precipitate$_{3}$
отношение	=	ratio$_{1}$	переменный	=	alternating$_{83}$
значение	=	value$_{1}$	отклонение	=	deflection$_{2}$
сплав	=	alloy$_{95}$	вооружение	=	arms$_{97}$
вид	=	species$_{4}$	напряжение	=	voltage$_{83}$

Undoubtedly the only recourse in these cases is to remove the subscript numbers and treat such alternatives as non-technical terms. The selection or rejection of such alternatives would necessarily be based on a much more sophisticated semantic classification than a synoptic chart of fields of science and technology. Fifty-five incorrect rejections out of the total 180 (30.5%) could be avoided by removing the subscript numbers. This procedure would not relieve the original ambiguity, but it would prevent loss of essential alter-

tives.

It should be observed that the largest number of examples were originally assigned the subscript number for mathematics. The necessity of removing this number in many cases is entirely in keeping with the widespread use of basic mathematical terms in the other sciences.

Faulty assignment of the contents of text passages to fields of science occurred in only two instances and gave rise to a very limited number of incorrect rejections. In the first instance, text passage 78 discussing the emergency release of landing gear, landing flaps and other assemblies were assigned to 91 (machinery, mechanisms, tools) and to 86 (control). In retrospect it seems obvious that 84 (aeronautical, acoustic) should have been used instead of 86 because appropriate alternatives for four semantic units, посадочный = landing$_{84}$, шасси = landing-gear$_{84}$, щитки = flaps$_{84}$, and полетный = gross$_{84}$, (10 occurrences in all) were incorrectly rejected. No new incorrect rejections would have been created by the substitution of 84 for 86. In the second instance, text passage 84 containing a discussion of sliver-lap machines was assigned only to 92 (production and manufacturing methods). Limiting the assignment in this way and not including 98 (textiles and paper) caused the incorrect rejection of лента = sliver-lap$_{98}$ and утолщенный = slubbed$_{98}$ (12 occurrences in all). The addition of number 98, while reviewing 12 incorrect rejections, would give rise to one other incorrect rejection, viz., гладкой = plain$_{98}$. It is very likely that a reconsideration of plain$_{98}$ (referring to the texture of cloth) would reveal that the subscript number should be removed entirely since it would be infrequently used in non-technical language. Twenty-two out of 180 (12%) incorrect rejections would thus be removed by adjusting the field to which the article was assigned.

The other half of the dichotomy of incorrect decisions is made up of incorrect selections. There were 176 incorrect selections; so the incorrect decisions were practically equally shared by rejections and selections. All incorrect selections have one particularly interesting feature: this is the only case where there is competition between alternatives with and without subscript numbers that cannot be solved. A few examples will illustrate this feature. The semantic unit вид has the equivalent "view/shape/species$_4$/aspect$_{82}$". Obviously, the third alternative "species" is found most often in articles on biology, and the fourth alternative in articles on linguistics. The first two alternatives "view" and "shape" may appear in all kinds of articles, however, including those on biology and linguistics. If вид were to appear in an article on biology, "view" and "shape" would be automatically rejected and "species" selected even though "view" or "species" might be the correct alternative. The imperfective infinitive приводить has the equivalent "(to)bring/cite/reduce$_1$". The alternative "reduce" is very likely to occur in any article on mathematics; but the other alternatives, "bring" and "cite", have a wide distribution in general technical and nontechnical literature and may easily be correct choices in the field of mathematics. The target-language equivalent of the adjective-substantive рабочую is "working/worker/operating$_9$". The alternative "operating" will frequently be appropriate in the area of technology, but even here the more generally applicable alternatives "working" and "worker" will be required. They must be retained therefore.

Because of the nature of the equivalents, alternatives with subscripts are always in competition with alternatives without subscripts. It is conceivable that alternatives with subscripts might also compete against other alternatives with subscripts. This could happen if a given target-language equivalent had either two or more alternatives bearing the same subscript number or two or more alternatives bearing two or more subscript numbers also assigned to the article being translated. Neither one of these two conditions obtained among the incorrect selections under discussion.

The only suggestion for remedying such competition between alternatives with and without subscript numbers is to eliminate the competition. That is, the subscript numbers should constitute a basis for selection or rejection of those alternatives that have subscript numbers. For example, in an article on linguistics "aspect" would be selected and "species" rejected, while "view" and "shape" would be retained. These latter alternatives still compete with "aspect" but not actively. Such treatment will remove all incorrect selections. The final resolution of semantic ambiguity would have to be made by more sophisticated procedures.

The results of this experiment in the automatic rejection or selection of technical terms are definitely encouraging. It is undoubtedly true, however, that another set of articles would reveal an almost entirely different set of incorrect selections and rejections. The process of first matching subscript numbers against the fields represented by different sets of articles and then evaluating the incorrect selections and rejections could possibly be repeated until a very high degree of refinement of subscript numbers is attained. Or better still, this procedure should be carried out in only one field of science at a time until the yield of incorrect selections and rejections is almost negligible. The Synoptic Chart of Fields and Subfields of Science and Technology is in no way regarded as a panacea for all the ills of semantic ambiguities among technical terms. A careful reevaluation and reconstruction of the chart may be indicated, but even this would not solve all problems of ambiguity. This chart, however, may very possibly constitute the basic framework for a more thorough-going semantic analysis in science and technology.