

TF-IDF results for Children's Drama

In reviewing the list of the 100 most important words for the children's corpus, I identified four of the most obvious main themes (items 1-4), but, of course, some words are difficult to assign unambiguously to any thematic group (item 5). For representativeness, I have colored words from different thematic groups with different colors:

1) Institution (**purple**): this group includes all vocabulary related to activities in a pioneer cell, a school, or, in some cases, a military school or orphanage;

2) Family (**blue**): this group includes all words naming family members and words closely associated with family;

3) Leisure, hobbies, and friendship (**green**): this group includes words that depict children's games and activities that are not related to school and pioneer activities, nor the family. We have collected such words into this thematic group;

4) Fighting with enemies (**yellow**): words that refer to the exposure of spies and pests, as well as vocabulary from conversations with antagonists, and torturers of children. Most often, these are words from plays where children suffer in the West. I attributed such words as “negro” to the same group because it is used in a derogatory form by white Americans in plays about a decaying life in America;

5) Neutral words (**grey**): words describing everyday objects and everyday actions, as well as words that thematically fit into more than two groups.

I assigned 13 words to two different groups. For example, the words *work* (noun) and *work* (verb) were allocated to the groups ‘Family’ and ‘State organizations’ at the same time. *Work* is closely embedded in both contexts, family and governmental organizations, so these words were placed in two thematic groups. The word *dacha* fell into two categories: ‘Family’ and ‘Leisure’. *Dacha* is closely connected with family and family traditions. A *dacha* plot belongs to the family and is an important place for spending time together. However, in children's plays, the *dacha* is also presented as a place where a child can play, meet peers, and go fishing – a place associated with a ‘happy childhood’ and carefree free time. I put the words *book*, *poem*, *caricature*, and *nut* into the categories of ‘State organizations’ and ‘Leisure’ at the same time. The characters in the plays read more often to study – to conform to the ideals of their state and the requirements of the school and pioneer cell. However, some characters read books and poems for pleasure rather than to do well in school. The same goes for caricatures – in children's plays, caricatures are drawn by characters both for their enjoyment and on assignment for the school wall newspaper. The collection and cultivation of *nuts* also has two contexts: for the prosperity of the collective farm and for pleasure, to eat for yourself. The words *commander*, *report*, and *battle* fall into the categories ‘Fighting the enemy’ and ‘Leisure’ because they are used in two contexts: the first context is direct when talking about military action, and the second context is children's war games when children imagine

themselves as soldiers on the battlefield. I put the words *captain*, *colonel*, and *officer* in the categories of ‘Fighting Enemies’ and ‘State Organizations’. Characters who wear these military titles are involved in military actions in plays about war. However, there are plays in the corps where the action takes place after the war and the colonels and officers act as teachers in a state military school (for example, in the play *Suvorovites* by Motorin and Ryss).

№	Lemma (in Russian)	Lemma (in English)	TF-IDF score
1	мистер	mister	608,2965
2	ребята	guys	496,3069
3	школа	school	410,9762
4	мисс	miss	262,2728
5	пионер	pioneer (n.)	235,7038
6	класс	class	231,1741
7	девочка	girl	228,7278
8	учиться	study (v.)	225,4935
9	мама	mother	219,9667
10	мальчик	boy	189,5873
11	папа	dad	185,9178
12	нужно	need	164,6987
13	урок	lesson	142,2244
14	пионерский	pioneer (adj.)	140,5312
15	отряд	squad	134,4635
16	работать	work (v.)	133,7486
17	маманька	maman	128,8008
18	двойка	grade F	125,6264
19	четверка	grade B	118,8931
20	господин	lord	118,2736
21	работа	work(n.)	116,0629
22	колхоз	collective farm	107,2686
23	капитан	captain	105,0092
24	пещера	cave	104,3775
25	мэр	mayor	102,3801
26	звено	link	101,5634
27	ой	oh	100,5878
28	посмотреть	look	92,85028
29	подождать	wait	88,42884

30	виноватый	guilty	87,32348
31	арифметика	arithmetic	85,95494
32	суворовец	suvorov cadet	82,56463
33	бабушка	grandmother	77,88477
34	получаться	manage/equal	75,8349
35	приют	shelter	74,73534
36	книжка	book	74,61176
37	дядя	uncle	73,38861
38	пятерка	grade A	73,06426
39	тройка	grade C	72,80533
40	директор	school principal	71,90377
41	пшеница	wheat	70,45482
42	орех	nut	69,49798
43	комсомолец	komsomolets	69,41903
44	дедушка	grandfather	69,19042
45	газета	newspaper	68,53235
46	дружба	friendship	68,49604
47	негр	negro	67,84539
48	галстук	tie	67,2729
49	здорово	great	67,2729
50	каток	ice rink	66,0517
51	эх	eh	65,21627
52	какой-то	some kind of	63,00555
53	сэр	sir	62,64645
54	обязательно	necessarily	61,90019
55	парта	desk	61,3213
56	вообще	at all	60,79483
57	поговорить	talk	60,79483
58	немец	German (n.)	59,93403
59	стих	poem	59,93403
60	дружить	be friends	59,6937
61	сначала	at first	59,68947
62	капитан-лейтенант	captain-lieutenant	59,50727
63	учитель	teacher	58,58411
64	чай	tea	58,58411
65	неужели	really	57,47875
66	па	Pa (abbr. "papa")	56,14395

67	полковник	colonel	54,267
68	вон	over there	53,0573
69	заходить	come in	53,0573
70	целый	whole	53,0573
71	лагерь	camp	52,91032
72	комсомол	Komsomol	52,8789
73	папаныка	daddy	52,84136
74	фронт	front	51,95194
75	катер	boat	51,73985
76	пропадать	disappear	50,84658
77	сорвать	disrupt	50,19697
78	деревня	village	50,14889
79	костер	bonfire	49,85725
80	шпион	spy	49,82356
81	дача	dacha	49,10127
82	выступать	perform	47,7026
83	исключать	exclude	47,7026
84	офицер	officer	47,7026
85	карикатура	caricature	46,96988
86	девчонка	girl (spoken)	46,47945
87	председатель	chairman	46,47945
88	командир	commander	46,42514
89	буржуин	bourgeois (n.)	46,23619
90	комсомольский	Komsomol (adj.)	45,32477
91	выполнять	fulfill	45,31978
92	тайна	secret	45,31978
93	ученик	student	45,31978
94	докладывать	report	45,25631
95	мел	chalk	44,36044
96	сельсовет	village council	44,36044
97	бой	battle	44,21442
98	собака	dog	44,21442
99	убегать	run away	44,21442
100	ясно	clear (adv.)	44,21442