

ЛИНГВОСТАТИСТИКА СЛОВООБРАЗОВАТЕЛЬНОЙ СИСТЕМЫ ЛЕКСИКО-СЕМАНТИЧЕСКОЙ ГРУППЫ «ЗРИТЕЛЬНОЕ ВОСПРИЯТИЕ»

О. Т. Косаренко

Воронежский государственный педагогический университет

С. В. Косаренко

Воронежский институт государственной противопожарной службы МЧС России

Поступила в редакцию 17 августа 2016 г.

Аннотация: статья посвящена лингвостатистическим особенностям организации словообразовательных гнезд глаголов зрительного восприятия по данным частотных словарей Национального корпуса русского языка. Рассмотрены параметры и законы распределения частот языковых единиц, характерные для текстов живой устной речи и художественной литературы.

Ключевые слова: лингвостатистический подход, словообразовательное гнездо, глагол, имя существительное, частотный словарь.

Abstract: the article focuses on the statistical study of verbs of visual perception and their derivatives. The study of word families is based on the data of frequency dictionaries of the Russian National Corpus. The article discusses parameters and frequency distribution of these linguistic units in colloquial speech and literary texts.

Key words: linguostatic, family of words, verb, noun, frequency dictionary.

Выработка методологии сопоставительного описания словообразовательных гнезд (СлГ) и изучения характера их взаимодействия сохраняет свою актуальность. В этом отношении статистический метод позволяет определить некие тренды, закономерности взаимосвязи (статистические связи или причинно-следственные зависимости) исследуемых объектов и признаков [1–3]. Предметом нашего исследования является направленность изменения частотных показателей употребительности элементов СлГ в текстах двух функциональных разновидностей – разговорной речи (РР) и художественной литературы (ХЛ). Материалом послужили лексические единицы (ЛЕ), входящие в СлГ, вершины которых образуют глаголы *видеть, глядеть, зреть* и *смотреть*, и достаточно полно описывающие смысловое поле «зрительное восприятие» (ЗВ).

Обратимся к двум информационно-справочным системам – Национальному корпусу русского языка (НКРЯ) и дополнительно к ресурсу Ngram Viewer в коллекции Google Books [4; 5]. Последний, предоставляя в наглядной графической форме данные по частоте встречаемости n-грамм для каждого года, дает несколько иную картину динамических параметров языковой эволюции, чем НКРЯ, поскольку поисковая система Google обрабатывает больший объем слов, но поиск заданных ЛЕ ведется по словоформам, а не по леммам и только в книгах, изданных до 2000 г.

Анализ четырех СлГ мы ограничили леммами, входящими в лексическое ядро языка, состав которого определили по Частотному словарю современного русского языка О. Н. Ляшевской, С. А. Шарова (2009 г.) [6], результирующему материалы НКРЯ. Так, согласно Новому частотному словарю русской лексики наиболее значимыми в дискурсе являются глаголы и существительные, представленные в СлГ с корнем *-вид-* шестью глаголами (*видать, видеть, видеться, ненавидеть, свидетельствовать, увидеть*) и двумя именами существительными (*вид, свидетель*); в СлГ с корнем *-гляд-* восьмью глаголами, из которых четыре глагола образуют видовую пару (*взглянуть, выгладеть, глядеть, глянуть, заглянуть/заглядывать, оглянуться/оглядываться, поглядеть/поглядывать, разглядеть/разглядывать*) и существительным *взгляд*; в СлГ с корнем *-зр-/-зер-/-зир-* глаголом *подозревать* и тремя именами существительными (*зеркало, зрение, зритель*); в СлГ с корнем *-смотр-* четырьмя глаголами (*посмотреть, предусмотреть, рассмотреть, смотреть*) и существительным *рассмотрение*. Следует отметить, что включение в СлГ с корнем *-зр-/-зер-* глаголов *подозревать, презирать* оправдано с диахронической точки зрения на процесс словообразовательной и семантической деривации. Гнезда разделили семантическое поле ЗВ и связанные (исторически) со зрительным восприятием психические процессы на полосы, участки разных очертаний. Пересечение различных линий-признаков, по которым идет сбли-

жение и отталкивание гнезд, образует само семантическое поле. Конкуренция наблюдается между всеми СлГ, системно взаимосвязанными, имеющими структурные и функциональные эквиваленты, и сближающимися по определенным признакам (например, по степени разветвленности, по распределению деривационной нагрузки по шагам деривации и т.п.), в том числе по количественным характеристикам.

Тот или иной ранг словесной единицы в частотном списке не может быть случайным. Этот ранг в то же время означает местоположение лингвистического объекта среди других, связанных с ним объектов. Частотный словарь, несмотря на то, что ограничивает словообразовательное пространство гнезд и исследуемое семантическое пространство, является репрезентативной выборкой, применимой к исходной совокупности (потенциалу гнезд), и дает достаточно хорошее представление о функциональной нагрузке (ФН) на СлГ в дискурсе. Функциональная нагрузка (ФН) – это употребительность той или иной части речи, определенной ЛЕ в потоке речи [7, с. 42].

На ФН влияют 1) общее количество высокочастотных слов, 2) количество ЛЕ, общих для исследуемых функциональных разновидностей и образующих лексическое ядро, 3) степень разнообразия, неповторяемости ЛЕ, уникальных для каждого стиля. Высокочастотные единицы закреплены за каждой функционально-стилистической подсистемой, но, взятые вместе, в какой-то мере воспроизводят матричное строение массива производной лексики современного языка и как бы заполняют клетки этой матрицы [8, с. 18], воплощают действие механизма словообразования. Для каждого СлГ совокупность этих позиций различна, в чем находит отражение коммуникативный и деривационный потенциал ЛСГ группы ЗВ применительно ко всем частям речи.

Частотные словари живой устной речи и художественной литературы в сопоставительном аспекте

Представленную в словарях выборку подмножеств и их объединение – наиболее частотных лемм РР и ХЛ по словарным спискам – образуют кластер, состав элементов (позиций) в котором определяется количеством оригинальных лемм, используемых в обеих сферах употребления. Образованная таким образом самостоятельная единица является вполне репрезентативной для определения степени разнообразия каждого СлГ и характера взаимодействия СлГ на уровне текста. Поэтому измерением числа значимых для СлГ лемм к общему числу оригинальных лемм кластера того же СлГ оценивается *коэффициент типичности лемм* (КТЛ) того или иного СлГ.

Кластер четырех СлГ имеет 74 позиции и включает в себя глагольных лемм вдвое больше (40), чем

лемм имен существительных (20), что в процентном отношении составляет 81 % лексического спектра частотных словарей. Именно в текстах ХЛ обнаруживаем большее разнообразие в выборе языковых средств (36 глаголов, 17 существительных), чем в записях живой устной речи (22 глагола и 10 существительных). Каждая функциональная разновидность вырабатывает свои «правила» использования общих свойств ЗВ. Словообразовательная система чутко (в зависимости от сферы и жанра) реагирует на изменение плана содержания, что находит выражение в изменении интенсивности частоты и ФН. Эта способность есть проявление системной детерминанты русского языка [9, с. 11].

ХЛ, как правило, в полной мере осваивает возможности РР; более того, количество слов (глаголов и существительных, используемых в ХЛ) в блоке СлГ с корнем *-гляд-* и СлГ с корнем *-зр-/-зер-/-зир-* совпадает с общим количеством оригинальных слов данных кластеров. Исключение составляет блок СлГ с корнем *-смотр-*, в частотном списке которого больше выбор ЛЕ в РР, чем в ХЛ (табл. 1). Отметим, что выбор того или иного типа дистрибуции между СлГ зависит от качественных и количественных характеристик ЛЕ, входящих в структуру СлГ.

Резонно предположить, что количество слов в блоке и их частотное распределение должны быть как-то зависимы друг от друга. Распределение признаков (КТЛ, ФН и др.) среди гнезд не может происходить равномерно, поскольку в отличие от деривационных отношений, существующих в СлГ, выравнивание отношений между элементами гнезд в частотном словаре (а значит, и в речи) с необходимостью следует логике развития объективных связей. Тем самым реализуется естественная, а не языковая система взаимосвязи объектов мира, где есть взаимодействие (т.е. слабая или сильная корреляция СлГ) и столкновение (иначе, доминирование в конкуренции СлГ) интересов с переменным успехом и множеством компромиссов (т.е. происходит приспособление и кооперация СлГ, перераспределение существенных признаков). Лексический профиль частотного словаря, разная структура СлГ в кластере определяются связями между гнездами, их конфигурация изменяется, если изменяется содержательное наполнение признаков (КТЛ и ФН). Частотное распределение одного признака связано с частотным распределением другого. Взаимосвязь между любыми двумя СлГ в обязательном порядке осуществляется в рамках словообразовательного пространства всех гнезд как общего для них объекта.

Очевидно, что существующая системная взаимосвязанность четырех СлГ и коррелированность статистических данных представляет собой следствие распределения во времени ФН на ЛЕ различных СлГ,

Частотность лемм в словарях живой устной речи и художественной литературы

Позиция	Частотный словарь живой устной речи			Частотный словарь художественной литературы		
	Лемма	Часть речи	Частота (ipm)	Лемма	Часть речи	Частота (ipm)
1	–	–	–	Всматриваться	v	19,5
2	Досмотреть	v	10,1	–	–	–
3	Несмотря	adv	21,4	Несмотря	adv	91,9
4	–	–	–	Осматривать	v	19,4
5	–	–	–	Осмотреть	v	37,1
6	Посмотреть	v	1009,2	Посмотреть	v	540,2
7	Просмотр	s	10,1	–	–	–
8	Просмотреть	v	11,3	–	–	–
9	Рассматривать	v	36,5	Рассматривать	v	56,6
10	Рассмотреть	v	11,3	Рассмотреть	v	25,7
11	Смотреть	v	1540,9	Смотреть	v	933,8
12	Смотреться	v	39,1	–	–	–

результат повышения/понижения частотности употребления. На графике Ngram Viewer Google Books в виде кривых встречаемости наглядно представлен возрастающе-убывающий характер динамики ЛСГ от среза к срезу (рисунок).

Сравнение величины ФН на словоформы данных глаголов также подтверждает факты: бесспорным лидером является *видеть*; последний ранг частоты отводится *зреть*; на протяжении последних двухсот лет в соотношении конкурирующих форм *видеть* и *смотреть* ротации не происходит, хотя в отдельных словоформах разрыв в величинах ФН может увеличиваться, как, например, в формах *вижу* и *смотрю*.

Особый интерес представляют статистический анализ сочетаемости ЛЕ с элементами контекста, совмещение парадигматического и синтагматическо-

го подходов при измерения динамических показателей языковой эволюции высокочастотных ЛЕ. Поиск наиболее частотных форм в НКРЯ по триграммам дает выборку вхождений, которая при последующем сопоставлении с данными другой поисковой системы позволяет скорректировать результаты. Так, для глагола *видеть* НКРЯ определил такие условия реализации (по первым 10 позициям списка вхождений): *я вижу что, я не вижу; я не видел, никогда не видел, я видел как, никто не видел; мы видим что, как мы видели; можно было видеть*.

Конкурирующие глаголы *видеть* (*видел, видим, вижу*) и *смотреть* (*смотрел, смотрит, смотрела, смотрю*) предпочитают не только разные словоформы, они используют разные дистрибутивные модели: *смотрел* на него, он *смотрел* на, *смотрел* на нее,

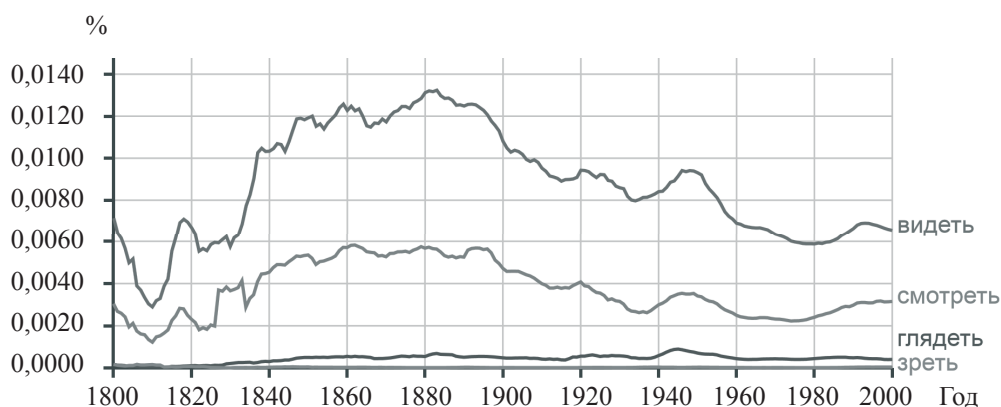


Рисунок. Тенденции использования фраз (n-грамм) в текстах с глаголами *видеть*, *глядеть*, *зреть*, *смотреть*

смотрел на меня, *я смотрел* на; *смотрит* на меня; она *смотрела* на, *я смотрю* на.

Среди словоформ 1-е место занимает форма *видел*, при этом в XXI в. наблюдается резкое изменение динамики роста: в 2000 г. частота равна 213 (принятый показатель частоты на миллион словоформ), а в 2014 г. – 895. Заметен также двукратный функциональный отрыв как от формы *видим*, так и от формы *вижу*, между которыми НКРЯ обнаруживает больший разрыв в 2014 г., чем в 2000 г. Эти формы на протяжении 14 лет также сохраняют динамику роста (в интервале с 31 и до 178 пунктов для формы *видим* и с 98 до 406 пунктов для формы *вижу*).

В целом в глагольной парадигме выделяем три группы словоформ, из которых все, кроме одной, формы прошедшего времени: в 1-й группе первенствует *видел*; затем со значительным отрывом идут формы *смотрел*, *посмотрел* и *выглядит*; в последней группе словоформа *глядел* резко выделяется на фоне глаголов *подозревал* и *презирал*, выражающих ментальное состояние, семантически опосредованно связанное со ЗВ.

Динамика языковых данных форм такова, что с начала Первой мировой войны наблюдается активизация почти всех форм ЛСГ (исключая словоформу *презирал*). НКРЯ фиксируют следующую динамику частот на 2014 г. при сопоставлении с 2000 г.: *видел* 895 – 213; *смотрел* 449 – 101; *посмотрел* 214 – 104; *выглядит* 270 – 45; *глядел* 83 – 18; *подозревал* 58 – 11; *презирал* 10,38 – 3,38. В предшествующем периоде находим изменение уровня ФН на ЛЕ *выглядит*: данная словоформа обошла формы *глядел*, *подозревал*, *презирал* в 60-е гг. XX в. Какая же причина ускорила этот рост? Статистика активности (сравнительно низкой для форм прошедшего времени и высокой – настоящего времени 1 л. и 3 л. глагола *выглядеть*) позволяет сделать предположение: то, как человек *выглядел*, менее значимо в речевой практике, чем то, как он *выглядит* сейчас; также важнее, как он/она *выглядит* (или они *выглядят*), чем самооценка внешности *выгляжу* (*выглядим*).

Специализация гнезд

Судя по представленным в табл. 2 данным, видно, что самая высокая степень реализации возможностей языковой системы наблюдается в части имен существительных у СЛГ корня *-зр-/-зер-/-зир-*, особенно в ХЛ (10 глаголов), а в отношении глаголов – у СЛГ корня *-гляд-* (в ХЛ 19 глаголов). Однако абсолютные величины дают не совсем верное представление о роли частей речи в современных текстах разной стилистической отнесенности, так как при равном количестве значимых глаголов в пределах одного СЛГ, например в СЛГ *-вид-* или *-смотр-* по семь глаголов и в РР, и в ХЛ, их доля относительно общего числа глаголов всех гнезд разная: в РР – 7/22, а в ХЛ – 7/36. Иначе говоря, глаголы наиболее активны именно в РР. Аналогично, пять и шесть существительных СЛГ *-вид-* составляют иные доли в РР (5/10) и в ХЛ (6/17).

Можно предположить, что если проявление исследуемого признака в одном СЛГ велико, то этот фактор сдерживает рост соответствующих показателей в других СЛГ и обуславливает *иное распределение* признака в объекте.

Следствием этого является распределение доли наиболее типичных для гнезд частей речи. Если лексический профиль позволяет сопоставить те СЛГ, которые имеют одинаковое количество ЛЕ в РР или ХЛ, то КТЛ, определяемый для каждого СЛГ, позволяет увидеть разброс поведения в изучаемом классе языковых единиц. Например, СЛГ с корнями *-вид-*, *-гляд-* и *-смотр-* имеют по семь глаголов (см. табл. 2), а разница КТЛ в гнездах иная от 0,29 до 0,35 и 0,58, соответственно.

Поскольку частотный словарь ограничивает словообразовательное пространство, то в матричной структуре кластеров всех СЛГ вес ЛЕ не может не зависеть от некоторых доминант, которые являются определяющими в данном словообразовательном пространстве в силу того, что они занимают большее, чем другие СЛГ, текстовое пространство. Воздействие одного СЛГ, доминирующего по какому-

Т а б л и ц а 2

Распределение наиболее частотных языковых единиц по частям речи

СЛГ	Частотный словарь живой устной речи				Частотный словарь художественной литературы			
	verb.	subs.	adj.	adv.	verb.	subs.	adj.	adv.
-вид-	7	5	1	4	7	6	2	3
-гляд-	7	1	–	–	19	1	–	–
-зир-	1	3	1	3	3	10	2	3
-смотр-	7	1	–	1	7	–	–	1
Всего	22	10	2	8	36	17	4	7

либо признаку, на другое распределяется неравномерно.

Так, доминирование глаголов СлГ с корнем *-гляд-* в кластере по числу отведенных данному грамматическому классу позиций (19 из 20, т.е. КТЛ равен 0,95 в СлГ и 0,47 всех занимаемых глаголами позиций), вероятно, сдерживает рост числа глаголов СлГ с корнем *-вид-* (8 или 0,33), и СлГ с корнем *-зр-/-зер-/-зир-* (8 или 0,16), и СлГ с корнем *-смотри-* (10 или 0,83).

Минимальный вес глаголов СлГ с корнем *-зр-/-зер-/-зир-* (КТЛ 0,16) восполняется за счет класса существительных (КТЛ 0,55) и других частей речи (КТЛ 0,27), причем имена существительные с этим корнем являются самыми востребованными в ХЛ среди всех СлГ.

Наиболее сбалансированный вес грамматических классов обнаруживает СлГ с корнем *-вид-* – КТЛ по 0,33. А в СлГ с корнем *-смотри-* части речи упорядочены менее гармонично со значительным перевесом в сторону глаголов. Примечательно, что суммарная частота 10 лемм несколько превосходит *ipm* 19 лемм СлГ с корнем *-гляд-* (табл. 3).

Интерпретация списков частотного словаря следующая. Судя по показателю *ipm* (принятая во многих словарях единица измерения; частота на млн слов корпуса), по частоте употребления глаголов в РР доминирует СлГ с корнем *-смотри-*, в ХЛ – СлГ с корнем *-вид-*, но среди имен существительных наиболее активно СлГ с корнем *-вид-*.

По частоте упоминания частей речи в дискурсе глаголы, как правило, превосходят все части речи, во всех стилях речи и всех СлГ. При учете *ipm* можно установить (см. табл. 3), какую именно ФН несет глагольная часть каждого СлГ. На примере СлГ с корнем *-зр-/-зер-/-зир-* в текстах ХЛ сдвиг ФН особенно заметен, так как нарушена пропорция между числом высокочастотных глаголов и их суммарными показателями *ipm*: в РР один глагол / *ipm* 10 и в ХЛ три глагола / *ipm* 105.

Аналогично значительное нарушение численных соотношений обнаруживаем в именной части СлГ. ФН существительных играет намного более заметную роль в ХЛ, чем в РР, например, для СлГ с корнем *-вид-* суммарная величина *ipm* может превосходить ожидаемый уровень в пять раз, а для СлГ с корнем *-гляд-* в 7,9 (при одинаковом количестве лемм частота (*ipm*) *взгляд* в РР = 55,4; в ХЛ = 439,5).

В целом при равном или незначительно большем количестве высокочастотных лемм в текстах разной стилевой принадлежности ФН растет именно в ХЛ.

ФН слабо коррелирует с меньшим/большим разнообразием словообразовательных типов того или иного СлГ. ФН распределяется следующим образом: суммарная частотность (*ipm*) 19 глаголов СлГ с корнем *-гляд-* почти равняется показателям семи глаголов СлГ с корнем *-смотри-*, занимающего ключевую позицию среди глаголов (ср.: 1515,6 и 1632,3). Богатство значимых для СлГ с корнем *-зр-/-зер-/-зир-* существительных не дает представление об их ФН в дискурсе: *ipm* 10 лемм (397,2) меньше не только *ipm* шести лемм СлГ с корнем *-вид-* (установленной нами доминанты), но и *ipm* одной леммы СлГ с корнем *-гляд-* (439,5).

Таким образом, лингвостатистический анализ словообразовательной системы лексико-семантической группы «зрительное восприятие» выявил согласованность данных и статистическую корреляцию между величинами, основанными на различных специальных признаках. Сравнимые СлГ обнаруживают больше сходства по одному из признаков, а в другом отношении могут далеко отстоять друг от друга. Стоит сказать, выбор того или иного типа дистрибуции между СлГ зависит от качественных и количественных характеристик ЛЕ, входящих в структуру СлГ. Глубинные причины тренда, в частности колебания частотности проявления речевого события с упоминанием способа и характера ЗВ, лежат за пределами языковой реальности, тем не менее рас-

Т а б л и ц а 3

Распределение частоты употребления глаголов и имен существительных (*ipm* и КТЛ)

СлГ	Общее количество оригинальных слов кластера	Суммарная частота частей речи (<i>ipm</i>)				Коэффициент типичности лемм	
		Словарь живой устной речи		Словарь художественной литературы			
		verb.	subs.	verb.	subs.	verb.	subs.
-вид-	24	1693	139	2097	721	0,29	0,25
-гляд-	20	238	55	1515	439	0,95	0,05
-зир-	18	10	84	105	397	0,16	0,55
-смотри-	12	2658	10	1632		0,58	0,08

смотренные в статье единицы измерения позволяют определять уровень исследуемых признаков в отношении СлГ.

Функциональное описание СлГ подтверждает выдвинутую гипотезу. От среза к срезу происходят колебания ФН вокруг внутрисистемного оптимума, поскольку лучшим решением проблемы является установление соответствия ресурсной базе СлГ. Одни СлГ и их отдельные ЛЕ укрепляют свою позицию за счет утративших некие «конкурентные преимущества» СлГ, и их ФН возрастает за счет снижения на ЛЕ других СлГ.

ЛИТЕРАТУРА

1. Частотный словарь национального корпуса русского языка : концепция и технология создания. – Режим доступа: <http://www.dialog-21.ru/digests/dialog2008/materials/html/53.htm>

2. Добрушина Н. Р. Как использовать Национальный корпус русского языка в образовании? / Н. Р. Добрушина

Воронежский государственный педагогический университет

Косаренко О. Т., доцент кафедры русского языка, современной русской и зарубежной литературы

E-mail: o.gordey@mail.ru

Тел.: 8 (473) 274-70-08

Воронежский институт государственной противопожарной службы МЧС России

Косаренко С. В., доцент кафедры иностранных языков и культуры речи

E-mail: irgies@rambler.ru

Тел.: 8-951-545-13-64

на // Национальный корпус русского языка : 2003–2005. Результаты и перспективы. – М., 2005. – С. 301–329.

3. Аверьянов Л. Я. Контент-анализ / Л. Я. Аверьянов. – М. : КноРус, 2007. – 456 с.

4. Национального корпуса русского языка. – Режим доступа: <http://ruscorpora.ru>

5. Google Books Ngram Viewer. – Mode of access: <https://books.google.com/ngrams/>

6. Ляшевская О. Н. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка) / О. Н. Ляшевская, С. А. Шаров. – М. : Азбуковник, 2009. – Режим доступа: <http://dict.ruslang.ru/freq.php>

7. Молчанова Л. В. Качественный и количественный аспекты лексико-семантической прогностики / Л. В. Молчанова. – Воронеж : ВГУ, 2007. – 173 с.

8. Земская Е. А. Словообразование как деятельность / Е. А. Земская. – М., 1992. – 221 с.

9. Богомазов Г. М. Современный русский литературный язык. Фонетика / Г. М. Богомазов. – М., 2001. – 352 с.

Voronezh State Pedagogical University

Kosarenko O. T., Associate Professor of the Russian Language, Modern Russian and Foreign Literature Department

E-mail: o.gordey@mail.ru

Tel.: 8 (473) 274-70-08

Voronezh Institute of State Firefighting Service of Russian Federation Ministry for Civil Defence, Emergencies and Elimination of Consequences of Natural Disasters

Kosarenko S. V., Associate Professor of the Foreign Languages and the Culture of Speech Department

E-mail: irgies@rambler.ru

Tel.: 8-951-545-13-64