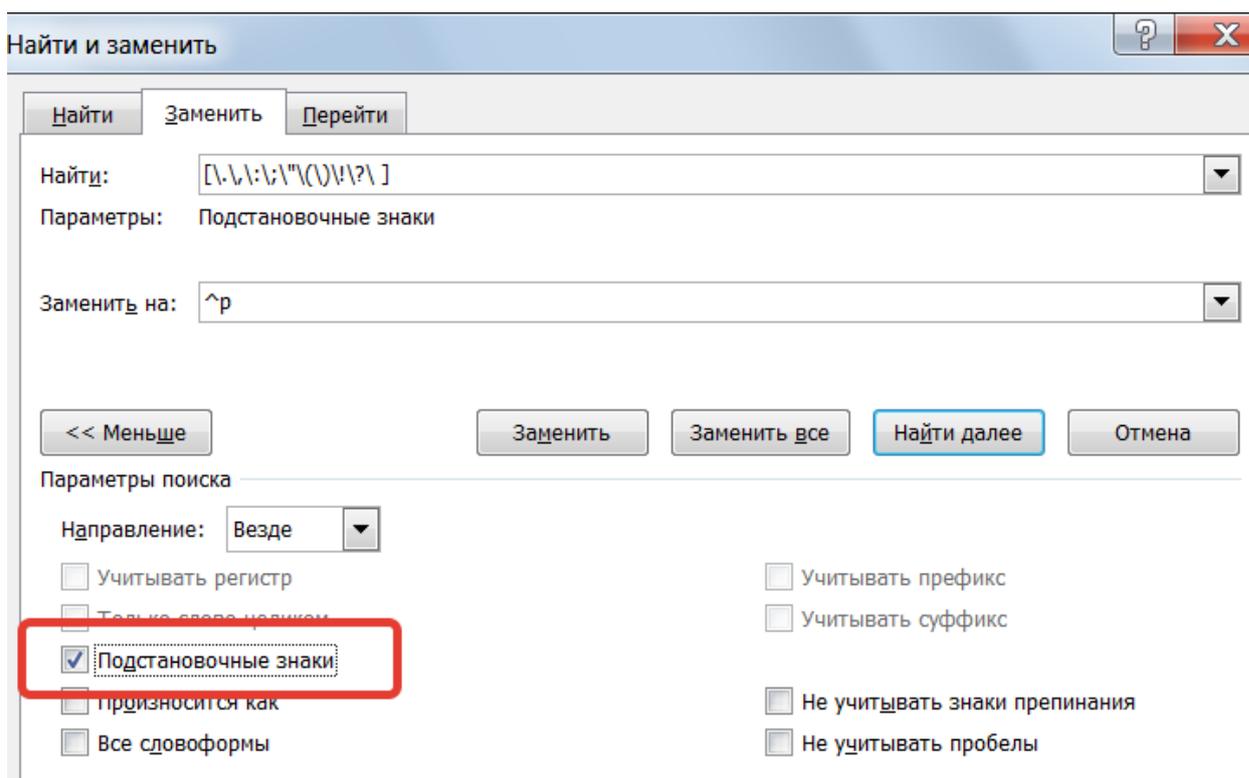


# Пошаговое руководство создания частотного словаря с помощью MS Word и MS Excel:

## 1. ПОДГОТОВКА ТЕКСТА В MS WORD:

**Шаг 1.** Копируем текст в MSWord, избавляемся от всех знаков пунктуации (в том числе от пробелов) и разносим каждое слово на отдельную строку с помощью инструмента **Заменить** (команда **Ctrl+H**):

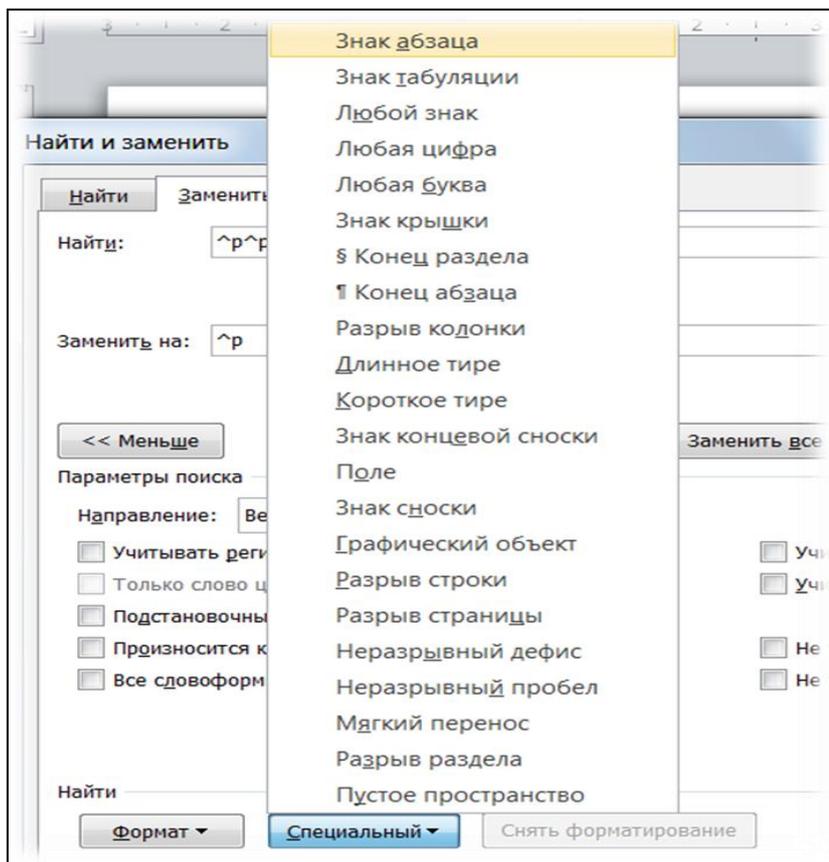


Для поиска/замены знаков препинания необходимо включить опцию «**Подстановочные знаки**» (включается по кнопке «**Больше >>**»). Не все знаки пунктуации можно перечислять «как есть», некоторые из них сами являются командой для функции подстановочных знаков, поэтому их нужно перечислять через обратный слеш \! или \?. Лучше поставить перед каждым знаком. Если нужно, скопируйте: `[\.\.\:\:\\"\\()\!?\ \-]`

Обратите внимание, что если оставить знак дефиса в этом перечне, то такие слова, как «когда-нибудь» будут разделены на два отдельных слова (как и предлагает Л.Н. Засорина в своем частотном словаре).

Замена производится на **Знак Абзаца** (кнопка «**Специальный**» или команда **^p**).

**Шаг 2.** После того, как знаки убраны, а каждое слово на отдельной строке, выделяем весь текст (команда **Ctrl+A**) и командой **Shift+F3** приведем все слова к единому регистру (текст получится более стройным, если все слова будут с маленькой буквы).



Затем следует избавиться от пустых строк с помощью того же инструмента замены (**Ctrl+H**), только теперь **Подстановочные знаки** должны быть **отключены**. Чтобы убрать пустые строки нужно **Найти** каждые два **Знака абзаца** (^p^p) и **Заменить** на один такой **Знак абзаца** (^p).

За один проход по кнопке **Заменить всё** количество пустых строк сократится только вдвое. Замену следует производить до тех пор, пока пустых строк совсем не останется.

Текст готов к дальнейшей обработке в MS Excel. Выделить весь текст **Ctrl+A**, вырезать **Ctrl+X**.

## 2. ОБРАБОТКА В MS EXCEL:

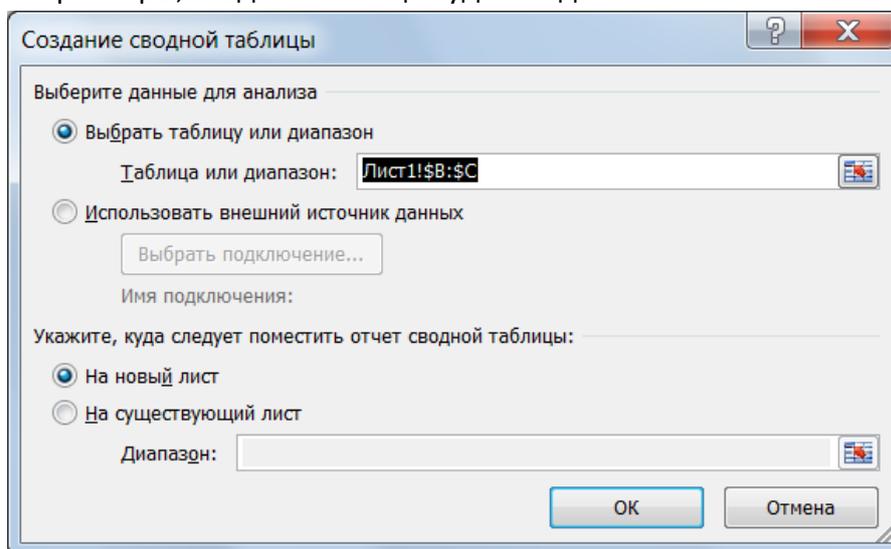
**Шаг 3.** Полученный список следует вставить в MS Excel в ячейки **A2** и **C2**. Первый столбец для **словоупотреблений**, второй для их **грамматических категорий**, а третий – слова, приведенные к их исходной форме (**леммы**). Назовите их соответственно:

	А	В	С
1	<b>Словоупотребления</b>	<b>Грам. категории</b>	<b>Леммы</b>
2	а	союз	а
3	бандероли	существительное	бандероль
4	бандероль	существительное	бандероль
5	беллетриста	существительное	беллетрист
6	близких	прилагательное	близкий
7	болезнью	существительное	болезнь
8	больной	прилагательное	больной
9	был	глагол	быть
10	был	глагол	быть
11	был	глагол	быть
12	было	глагол	быть
13	было	глагол	быть

**Шаг 4.** Лемматизация для столбца **С** и определение для каждого словоупотребления его грамматической категории в столбце **В** производится самостоятельно (при этом MS Excel помогает с автозаполнением слов, упомянутых в списке выше).

**Шаг 5.** Для создания **частотного словаря**, выделим столбцы с грамматическими категориями и леммами, проведя мышью по их буквам: **В** и **С**.

На вкладке **Вставка** слева выбираем **Сводная таблица**. Рекомендуется не менять предложенные параметры, сводная таблица будет создана на новом листе:



Для заполнения таблицы 1 из пояснения к Лабораторной работе № 3 включите поле Грамматических категорий и перетащите его в область **Σ Значения**. Затем в столбце **С** рассчитайте **относительную частоту** грамматических категорий.

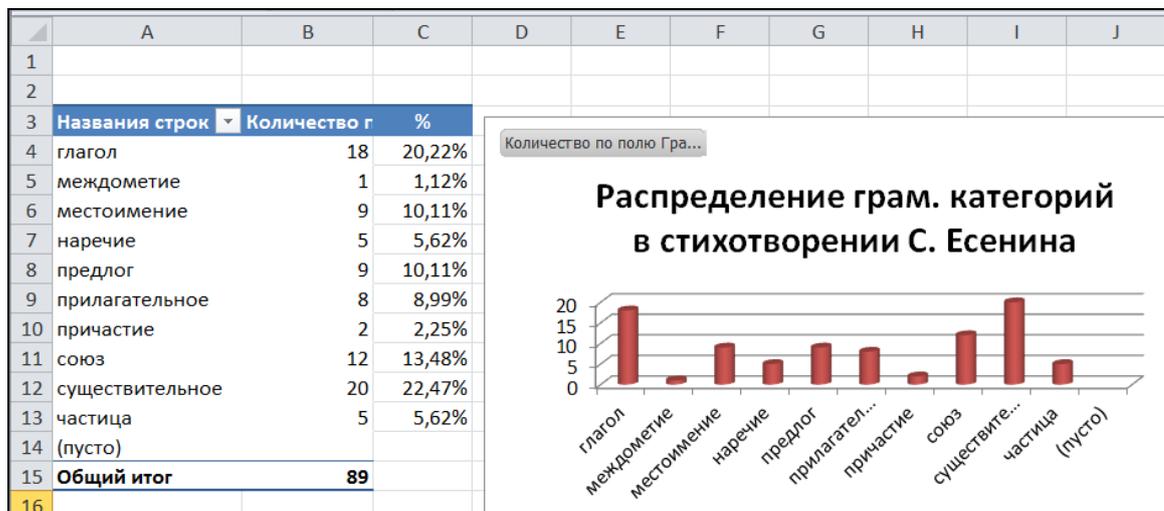
	A	B	C	D	E
1					
2					
3	Названия строк	Количество			
4	глагол	18			
5	междометие	1			
6	местоимение	9			
7	наречие	5			
8	предлог	9			
9	прилагательное	8			
10	причастие	2			
11	союз	12			
12	существительное	20			
13	частица	5			
14	(пусто)				
15	<b>Общий итог</b>	<b>89</b>			
16					
17					
18					

1. Включить Грамм. кат.

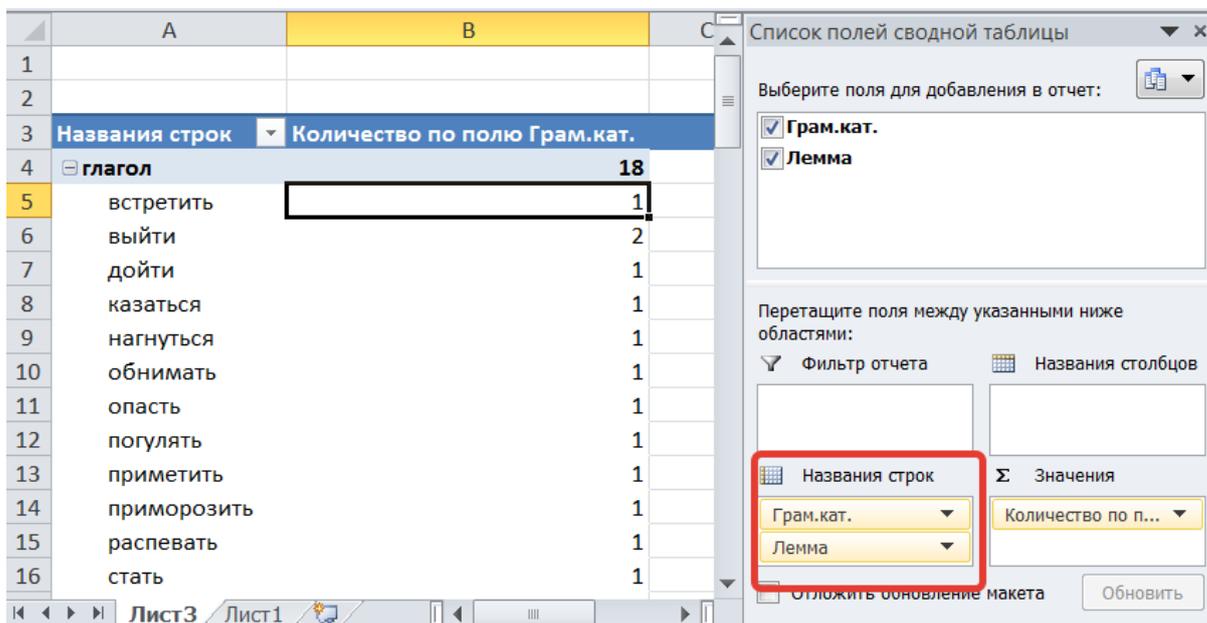
2. Перетащить Грамм. кат в поле Значения (появится абс. количество для каждой категории в столбце В).

3. В столбце С вычислите относительные частоты категорий

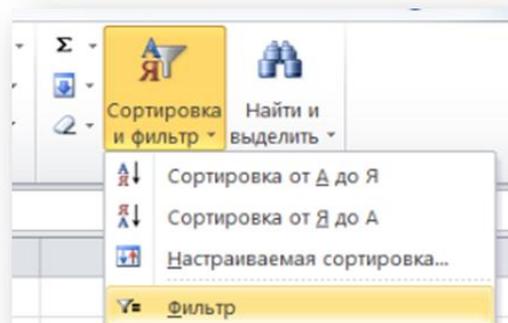
**Шаг 6.** Для наглядности простройте диаграмму из данных **A** и **C** (несмежные диапазоны выделяются с прижатой клавишей **Ctrl**). Диаграмму можно будет вставить в лабораторную работу для каждого из отрывков (см. рис. ниже).



**Шаг 7.** Для заполнения таблицы 2 из пояснения к Лабораторной работе № 3 включите опцию **Леммы** чтобы получить следующую картину (**Названия строк** должны стоять именно в таком порядке – **Грам. кат.** выше, чем **Лемма**):



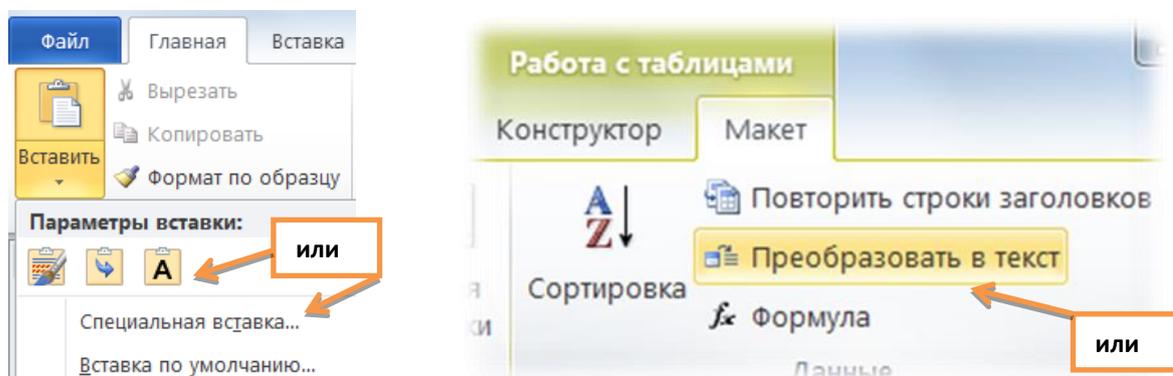
Чтобы отсортировать глаголы и другие части речи не по алфавиту, а по частоте использования, станьте на ячейку с первым значением (на рис. это ячейка B5) и нажмите сортировку (от Я до А). Кнопка сортировки находится на Главной вкладке справа (рядом с опцией **Найти и выделить**).



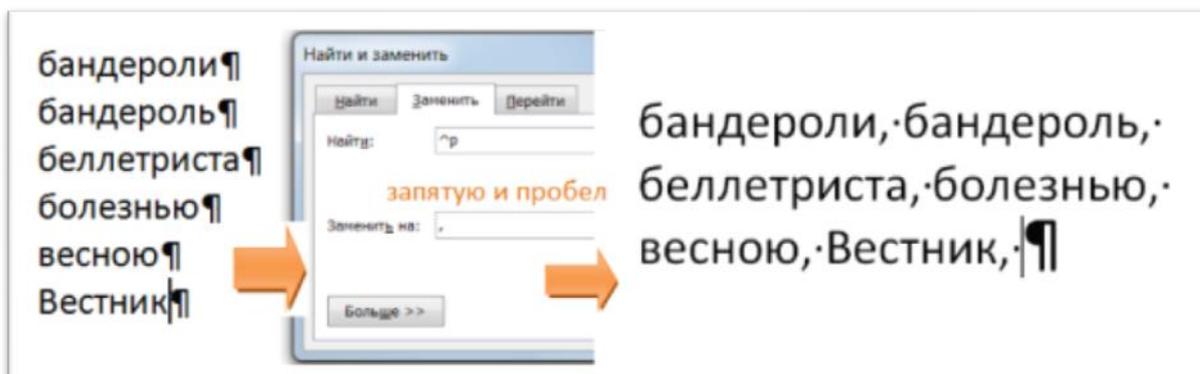
**Шаг 8.** Создайте таблицу 2 и заполните ее по следующему принципу: в колонку **Частота: >=2** следует копировать только те строки, в которых перечислены слова, упомянутые более 1 раза в произведении. В колонку Частота 1 – остальные, уникальные леммы.

Существительные		
Частота: >=2		Частота: 1
клён	3	береза, верба, деревня, дом, дорога, доска, жена, лето, нога, песня, попойка, скромность, сосна, сторож, сугроб
метель	2	
Глаголы		
Частота: >=2		Частота: 1
выйти	2	встретить, дойти, казаться, нагнуться, обнимать, погулять, приметить, приморозить, распевать, стать, стоять, увидеть, услышать, утонуть, утратить
опасть	2	
<i>и т.д.</i>		
Частота: >=2		Частота: 1

Если со вставкой в колонку **Частота: >=2** вопросов обычно не возникает, то при вставке уникальных лемм следует учитывать, что при копировании из MS Excel в ячейку таблицы MS Word вставляется вложенная таблица. Ее следует **преобразовать в текст** на вкладке **Макет**, либо изначально **вставлять** скопированный текст через **Специальную вставку**.



Отсортируйте строки уникальных лемм по алфавиту (от А до Я). Затем произведите замену Знаков Абзаца **^p** на **запятую** с пробелом:



**P.S.** Аналогичные операции можно проводить и **для построения частот букв в тексте** (например, для исследования механизма звукоцветовых ассоциаций по А.П. Журавлеву). Для этого в Шаге 1 вместо диапазона знаков пунктуации в поле «Найти» следует указать в специальных командах «**Любой знак**» (или **^?**), а в поле «Заменить на» выбрать «**Искомый текст**» (или **^&**) и «**Знак Абзаца**» (или **^p**). Опция «Подстановочные знаки» при этом должна быть **отключена**. Все знаки распределятся на отдельных строках, при сортировке по возрастанию верхние строки (до букв) можно будет легко удалить.