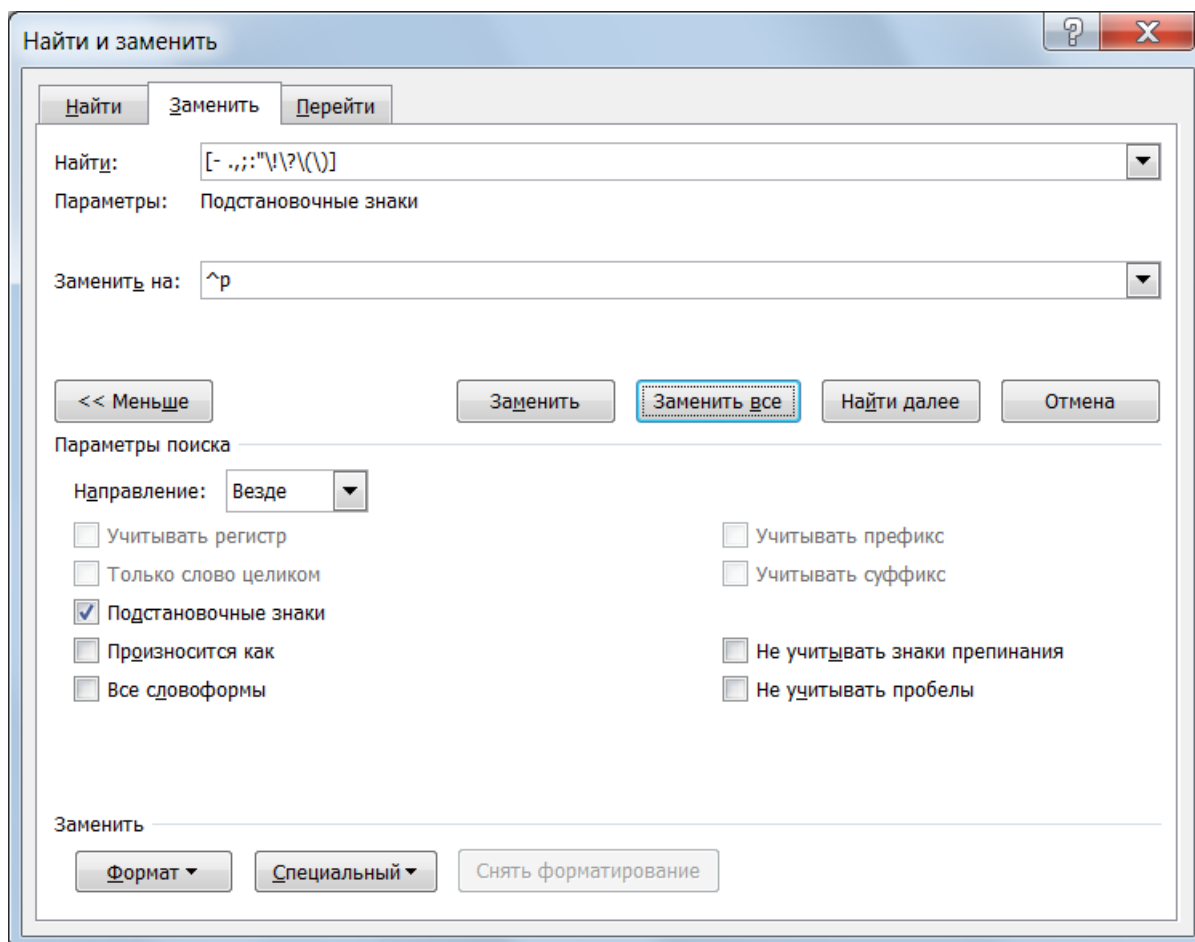


# Пошаговое руководство создания частотного словаря с помощью MS Word и MS Excel:

## 1. Подготовка текста в MS Word:

**Шаг 1.** Копируем текст в MS Word, избавляемся от всех знаков пунктуации и разносим каждое слово на отдельную строку:

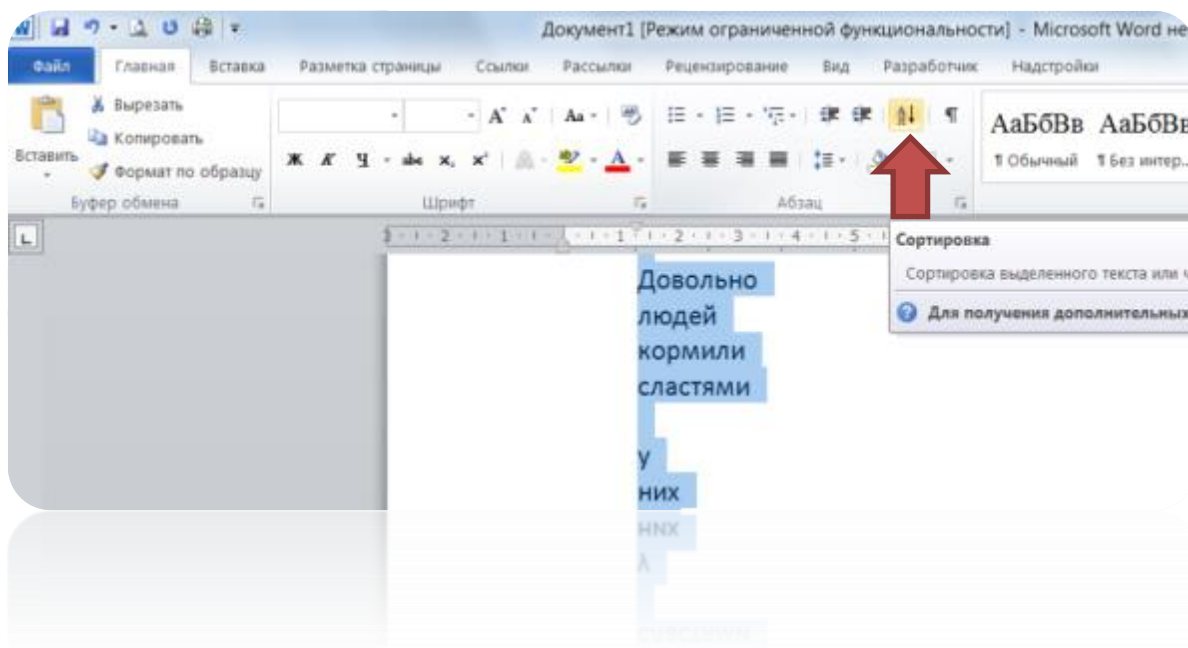


Для поиска/замены знаков препинания необходимо включить опцию «**Подстановочные знаки**» (включается по кнопке «**Больше >>**»). Не все знаки пунктуации можно перечислять «как есть», некоторые из них сами являются командой для функции подстановочных знаков, поэтому их нужно перечислять через обратный слеш \! или \?. 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).

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 **сортируем** по возрастанию. На этом этапе можно удалить сверху пустые строки (до первого значимого слова):



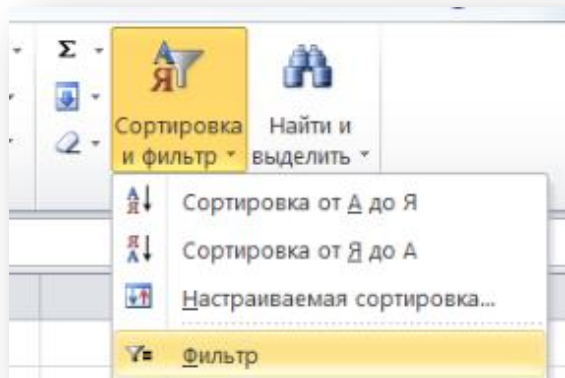
## 2. ОБРАБОТКА В MS EXCEL:

**Шаг 3.** Отсортированный список следует вставить в MS Excel, в ячейку **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).

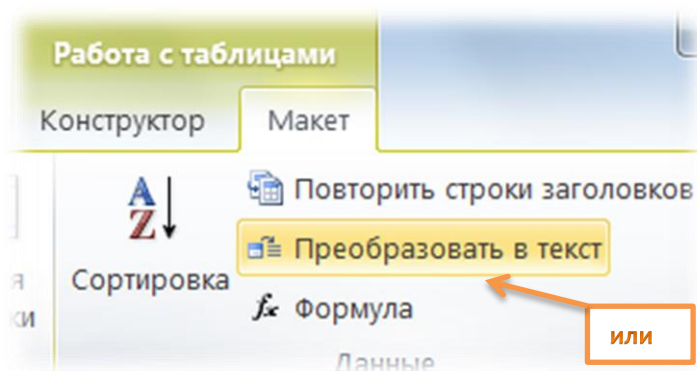
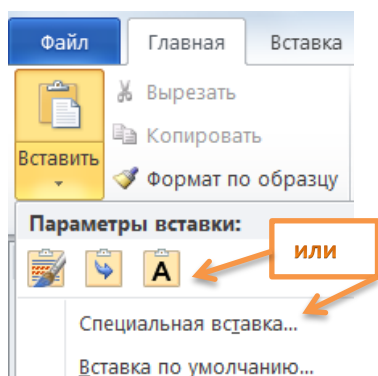
**Шаг 5.** Включить в первой строке фильтр для сортировки:



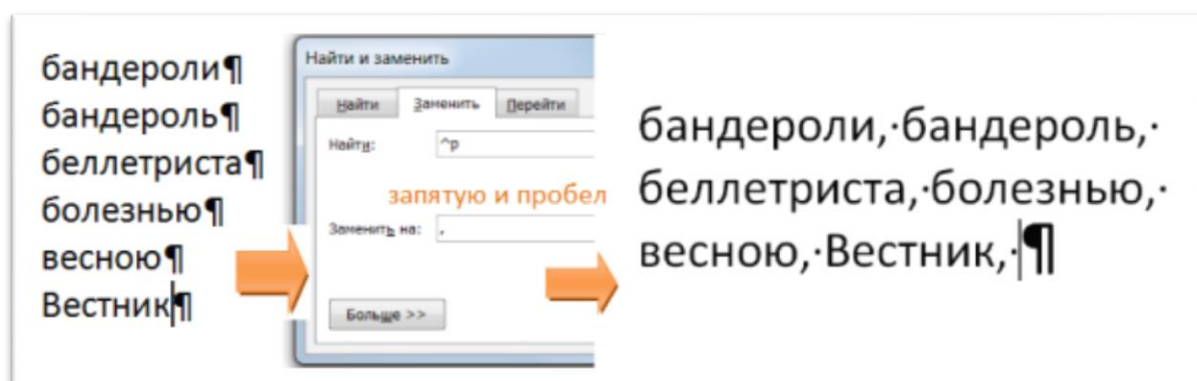
Это даст возможность фильтровать список по категориям и работать с каждой категорией отдельно:

	А	В	С
1	<b>Слова</b>	<b>Грам. категори</b>	<b>Леммы</b>
3	€	Поиск	бандероль
4	€	<input checked="" type="checkbox"/> (Выделить все)	бандероль
5	€	<input type="checkbox"/> глагол	беллетрист
7	€	<input type="checkbox"/> предлог	болезнь
26	Е	<input type="checkbox"/> прилагательное	весна
27	Е	<input checked="" type="checkbox"/> союз	Вестник
		<input checked="" type="checkbox"/> существительное	
		<input type="checkbox"/> (Пустые)	

**Шаг 6.** Для распечатки **словоупотреблений** в качестве приложения к лабораторной работе скопируем **столбец А** для каждой категории отдельно в MS Word. Используйте при этом Специальную вставку – **Неформатированный текст**, чтобы избавиться от табличных рамок, либо если вставили таблицей, то **преобразовать** ее в текст на вкладке Макет.

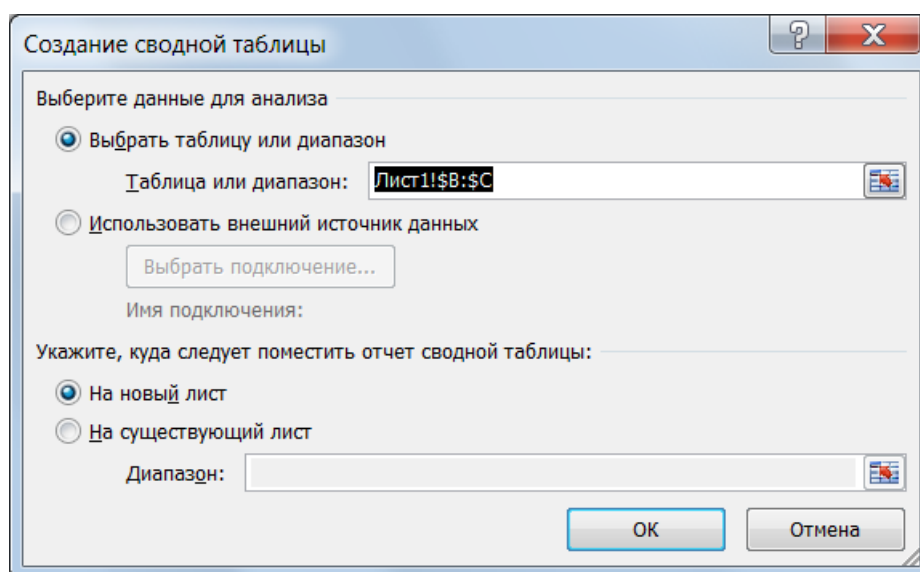


Затем производится обратная замена (Знак Абзаца ^p на **запятую** с пробелом – опция «Подстановочные знаки» при этом должна быть **отключена**).



В конечном итоге из столбца **А** (суммарно) должен получиться исходный текст, но распределенный на «кучки» – существительные, прилагательные и т.д.

**Шаг 7.** Для создания **частотного словаря**, выделяем столбцы **С** и **В** (с леммами и грамматическими категориями) и на вкладке **Вставка** слева выбираем **Сводная таблица**. Рекомендуется не менять предложенные параметры, сводная таблица будет создана на новом листе:

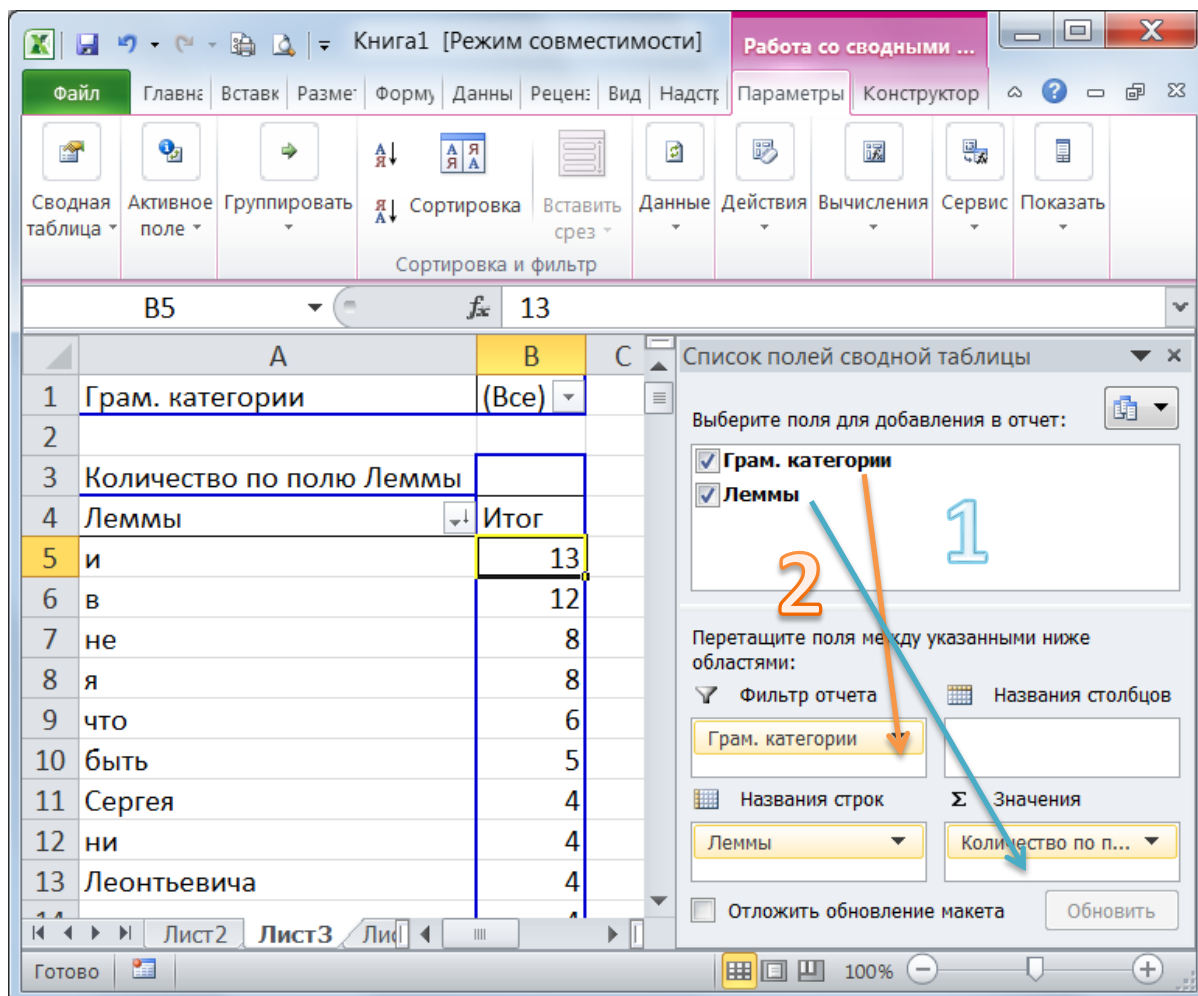


Опираясь на схему последнего рисунка выполняем следующие действия:

1. Справа в списке полей: **включить** поле **Леммы** (слева появятся уникальные слова из столбца Леммы), **перетащить** включенное поле вниз в поле **Значения** (с пиктограммой суммы  $\Sigma$  – появится количество для каждого из слов). **Отсортируйте Итог** по убыванию (от Я до А), чтобы большее количество употреблений было вверху.

**Частотный словарь готов.**

2. Теперь в правом верхнем поле можно включить еще поле **грамматических категорий** и перетащить в окно **Фильтр отчета** для отбора по категориям. Станет возможным отбор слов по категориям – в ячейке (Все) над словом Итог.



**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.