Создание частотного словаря словоформ с помощью инструментов Microsoft Word и Excel

Для работы вам понадобится русскоязычный текст объемом **около 3 стр**. (порядка 1200-1400 слов).

В ходе выполнения работы каждому студенту необходимо представить в отдельной папке:

1) исходный вариант текста-источника (в формате **Microsoft Word**);
2) трансформированный вариант исходного текста без знаков препинания в виде списка слов (в формате **Microsoft Word**);
3) Частотный словарь словоформ (в формате **Microsoft Excel**).

1. Первым этапом в построении частотного словаря станет удаление знаков препинания из текста. Для этого необходимо использовать опцию «Правка» → «Найти» → «Заменить».

В поле «Найти» пишем нужный знак препинания (например, запятую), в поле заменить – ставим пробел. (Рис. 1.)

Далее нажимаем на кнопку «Заменить все». (Рис. 2). Запятые будут убраны из всего текста.

Аналогичную операцию проводим с оставшимися знаками препинания (точки, кавычки и т.д.).

Рис. 1

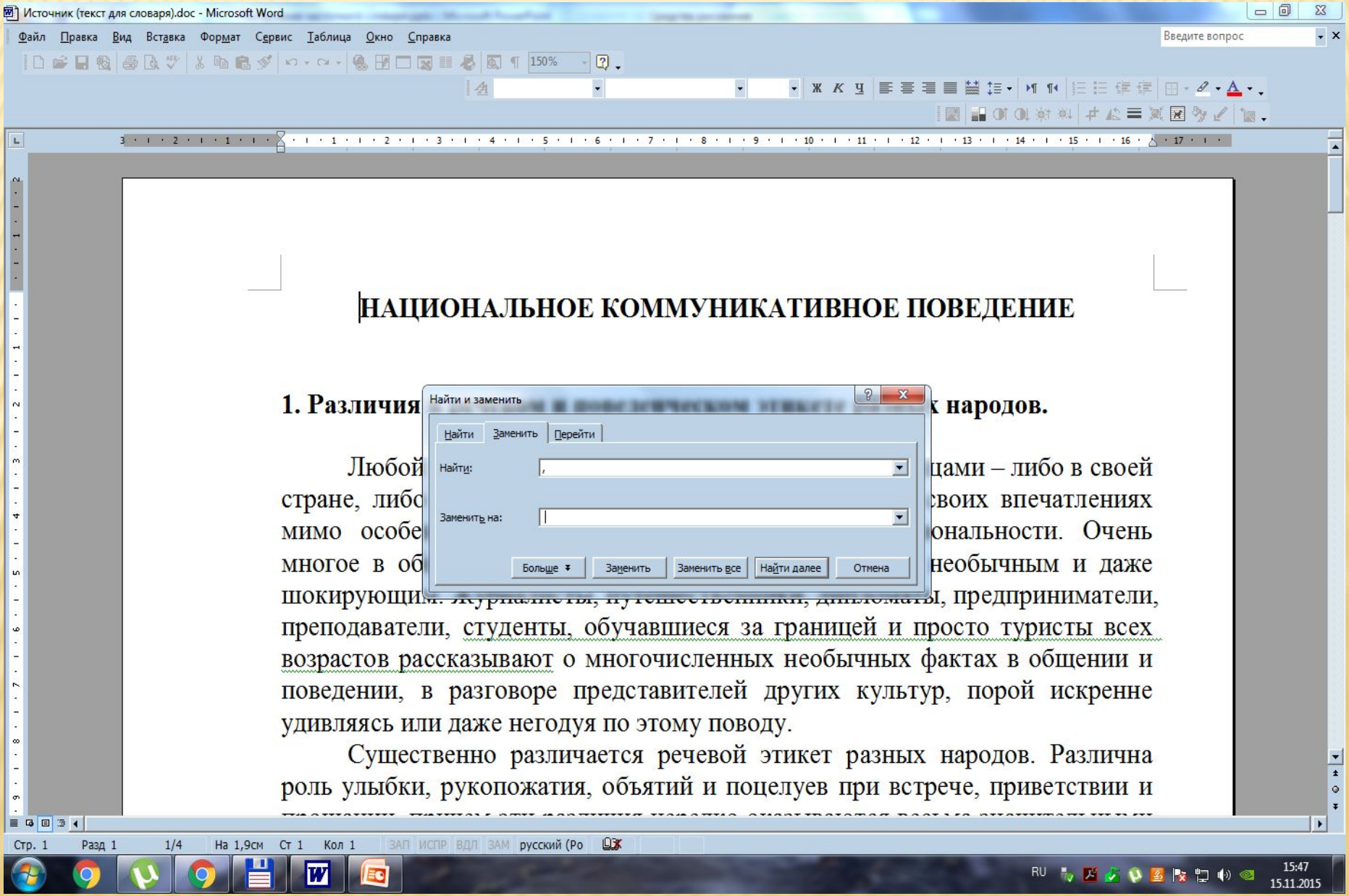
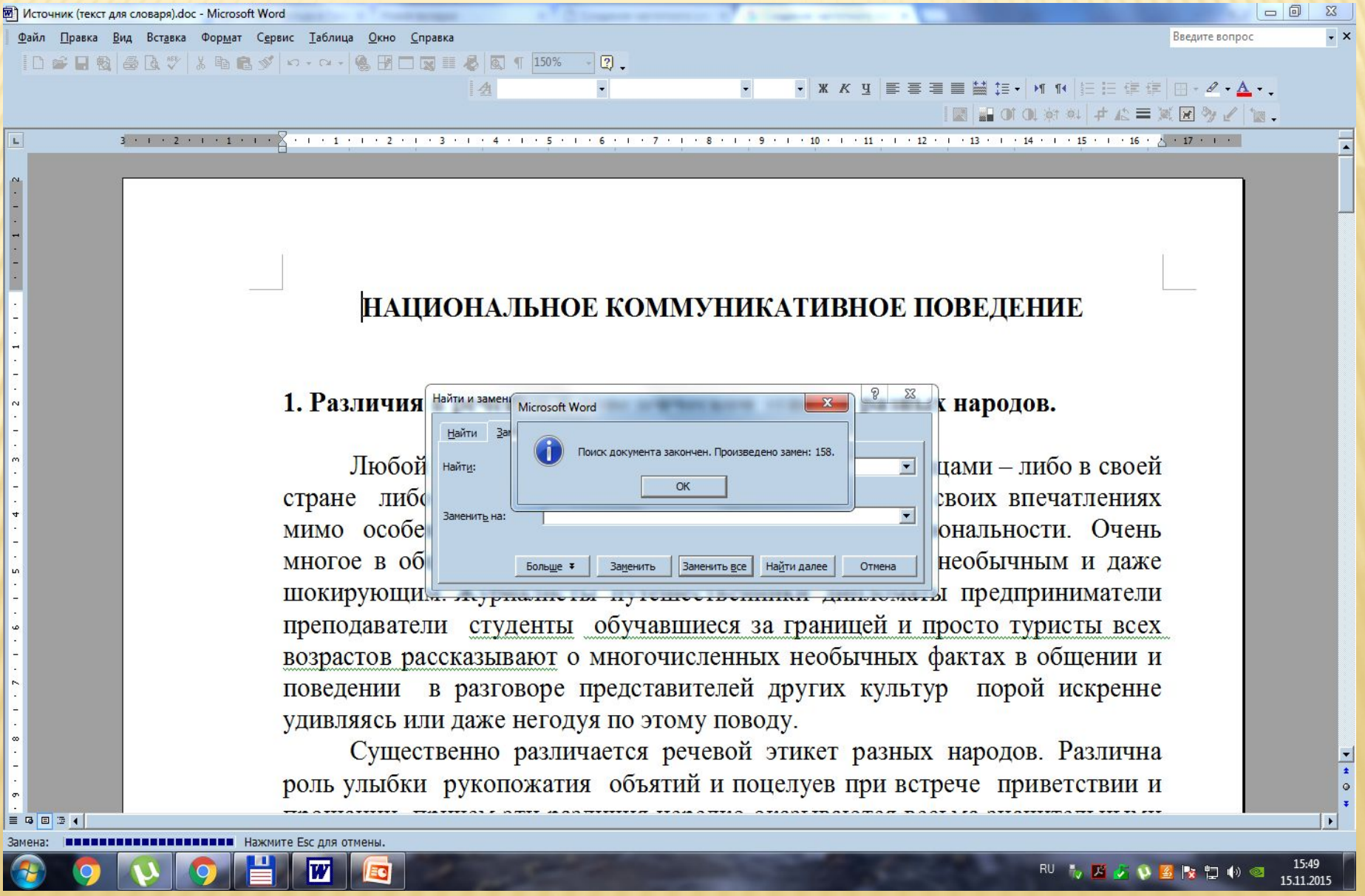

Источник (текст для словаря).doc - Microsoft Word
Файл Правка Вид Вставка Формат Сервис Таблица Окно Справка
Введите вопрос
НАЦИОНАЛЬНОЕ КОММУНИКАТИВНОЕ ПОВЕДЕНИЕ
1. Различия
народов.
Найти и заменить
Найти Заменить Перейти
Найти:
Заменить на:
Больше Заменить Заменить все Найти далее Отмена
Любой
цами – либо в своей
стране, либо
своих впечатлениях
мимо особе
ональности. Очень
многое в об
необычным и даже
шокирующим. Журналисты, путешественники, дипломаты, предприниматели,
преподаватели, студенты, обучавшиеся за границей и просто туристы всех
возрастов рассказывают о многочисленных необычных фактах в общении и
поведении, в разговоре представителей других культур, порой искренне
удивляясь или даже негодуя по этому поводу.
Существенно различается речевой этикет разных народов. Различна
роль улыбки, рукопожатия, объятий и поцелуев при встрече, приветствии и
Стр. 1 Разд 1 1/4 На 1,9см Ст 1 Кол 1 ЗАП ИСПР ВДЛ ЗАМ русский (Ро
RU
15:47
15.11.2015

Рис. 2.


Источник (текст для словаря).doc - Microsoft Word
Файл Правка Вид Вставка Формат Сервис Таблица Окно Справка
Введите вопрос
НАЦИОНАЛЬНОЕ КОММУНИКАТИВНОЕ ПОВЕДЕНИЕ
1. Различия
народов.
Любой
цами – либо в своей
стране либо
своих впечатлениях
мимо особе
ональности. Очень
многое в об
необычным и даже
шокирующим. Журналисты путешественники дипломаты предприниматели
преподаватели студенты обучавшиеся за границей и просто туристы всех
возрастов рассказывают о многочисленных необычных фактах в общении и
поведении в разговоре представителей других культур порой искренне
удивляясь или даже негодуя по этому поводу.
Существенно различается речевой этикет разных народов. Различна
роль улыбки рукопожатия объятий и поцелуев при встрече приветствии и
Найти и замен
Найти
Microsoft Word
Поиск документа закончен. Произведено замен: 158.
OK
Найти:
Заменить на:
Больше
Заменить
Заменить все
Найти далее
Отмена
Замена:
Нажмите Esc для отмены.
RU
15:49
15.11.2015

2. Следующий шаг – постановка всех слов текста в столбик. Для этого также будем использовать опцию «Правка» → «Найти» → «Заменить».

В поле «Найти» ставим знак пробела (нажать один раз), в поле заменить нам нужно указать знак абзаца. (Рис. 3.) Для этого в поле «Специальные знаки» выберем параметр «Знак абзаца».

Далее нажимаем на кнопку «Заменить все». (Рис. 4). Слова должны расположиться друг под другом.

Рис. 3

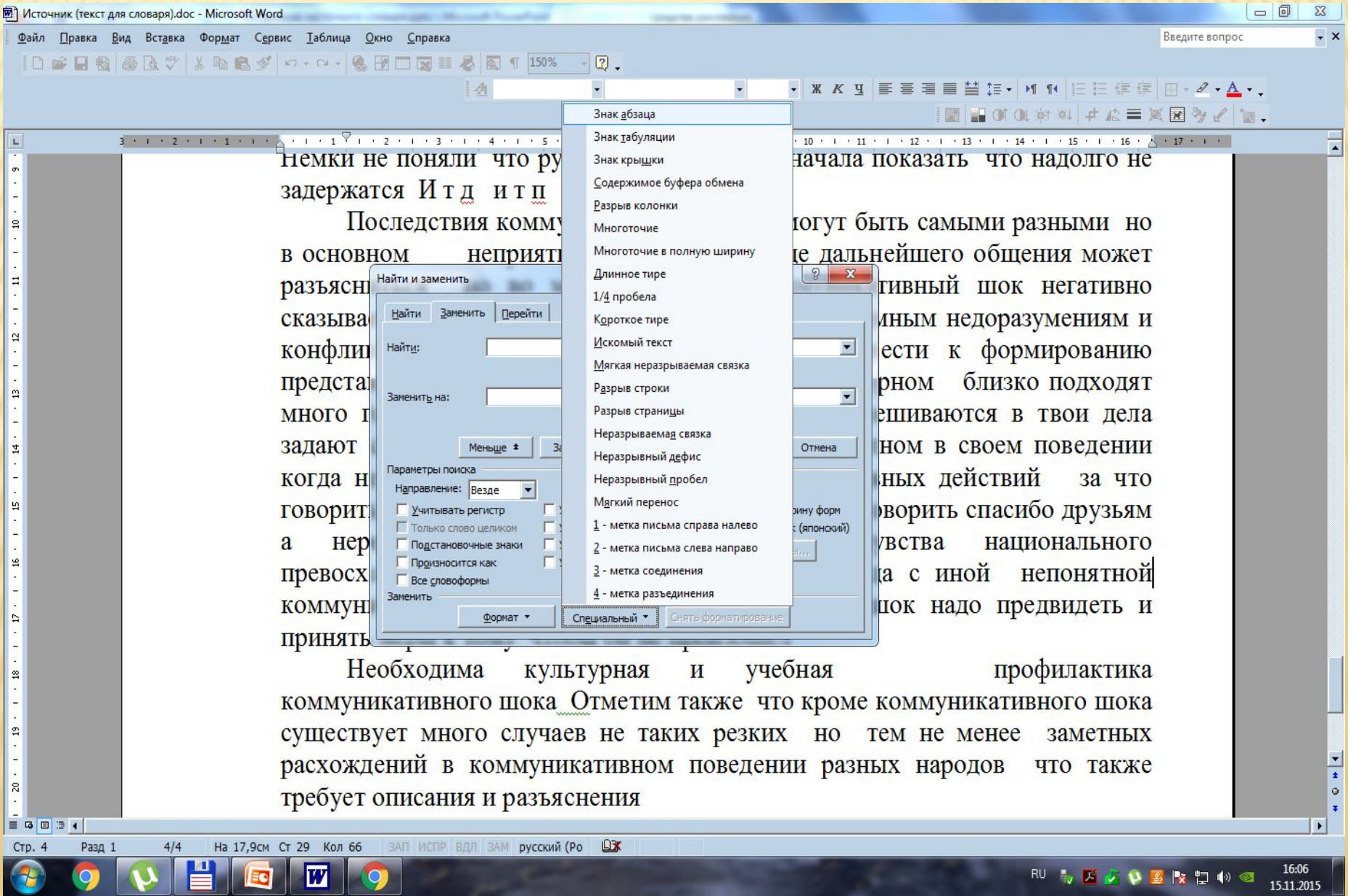
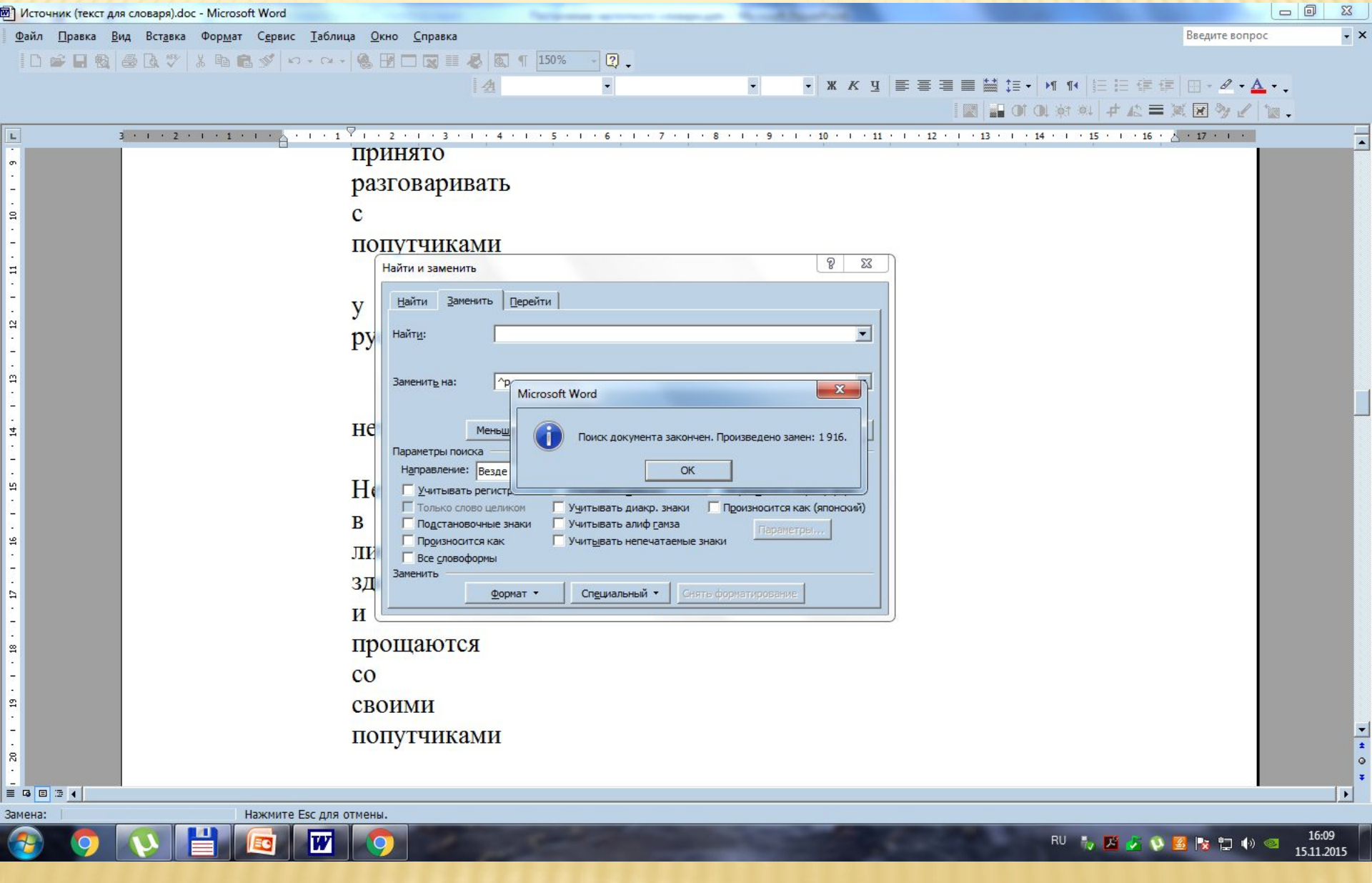
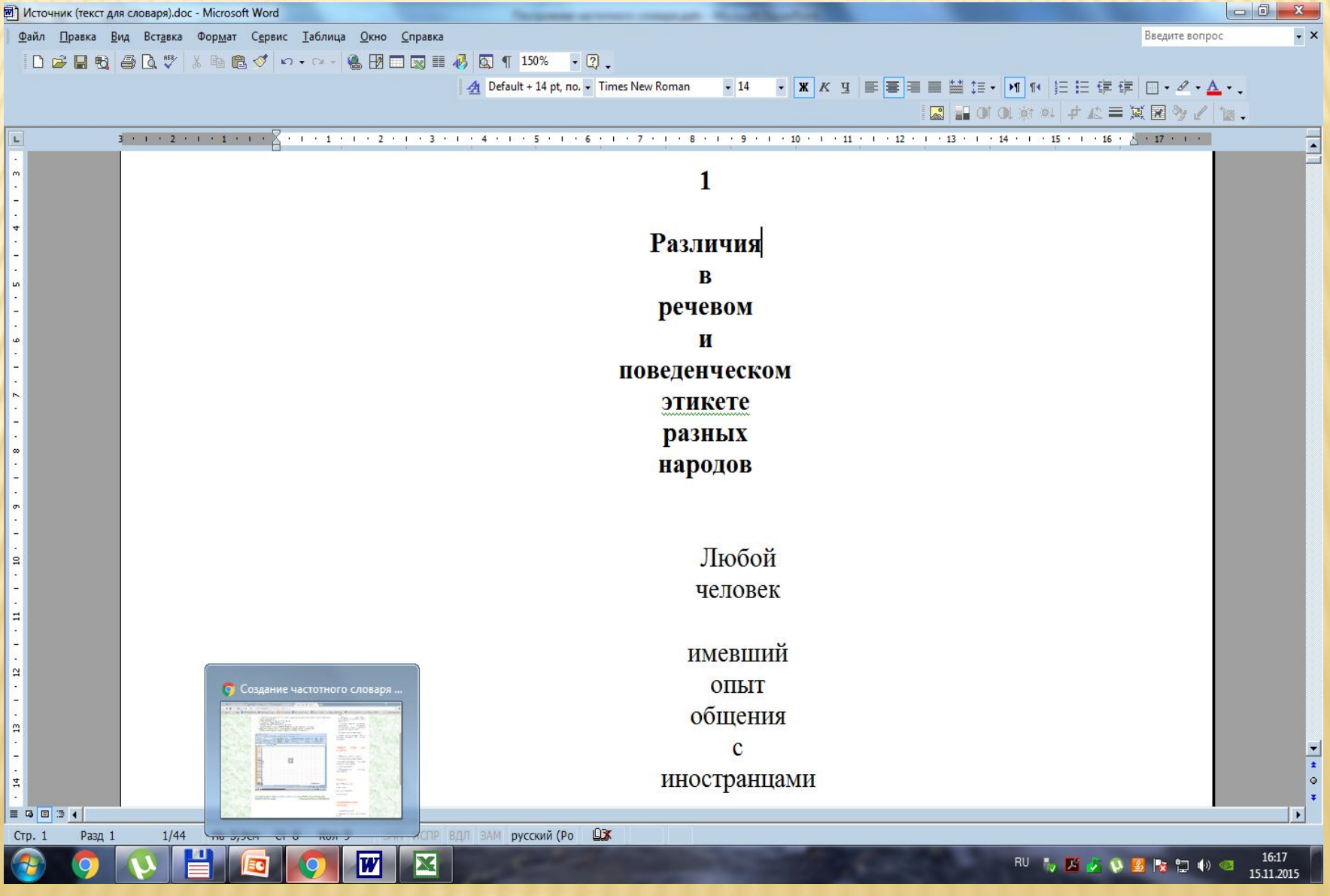

Источник (текст для словаря).doc - Microsoft Word
Файл Правка Вид Вставка Формат Сервис Таблица Окно Справка
Введите вопрос
Знак абзаца
Знак табуляции
Знак крышки
Содержимое буфера обмена
Разрыв колонки
Многоточие
Многоточие в полную ширину
Длинное тире
1/4 пробела
Короткое тире
Искомый текст
Мягкая неразрываемая связка
Разрыв строки
Разрыв страницы
Неразрываемая связка
Неразрывный дефис
Неразрывный пробел
Мягкий перенос
1 - метка письма справа налево
2 - метка письма слева направо
3 - метка соединения
4 - метка разъединения
Найти и заменить
Найти
Заменить
Перейти
Найти:
Заменить на:
Меньше
Параметры поиска
Направление: Везде
Учитывать регистр
Только слово целиком
Подстановочные знаки
Произносится как
Все словоформы
Заменить
Формат
Специальный
Снять форматирование
Отмена
Немки не поняли что ру начала показать что надолго не
задержатся И т д и т п
Последствия комму могут быть самыми разными но
в основном неприяти е дальнейшего общения может
разъясн тивный шок негативно
сказыва ным недоразумениям и
конфли ести к формированию
предста рном близко подходят
много п ешиваются в твои дела
задают ном в своем поведении
когда н ных действий за что
говорит ворить спасибо друзьям
а нер вства национального
превосх а с иной непонятной
коммун шок надо предвидеть и
принять
Необходима культурная и учебная профилактика
коммуникативного шока Отметим также что кроме коммуникативного шока
существует много случаев не таких резких но тем не менее заметных
расхождений в коммуникативном поведении разных народов что также
требует описания и разъяснения
Стр. 4 Разд 1 4/4 На 17,9см Ст 29 Кол 66 ЗАП ИСПР ВДЛ ЗАМ русский (Ро
RU
16:06
15.11.2015

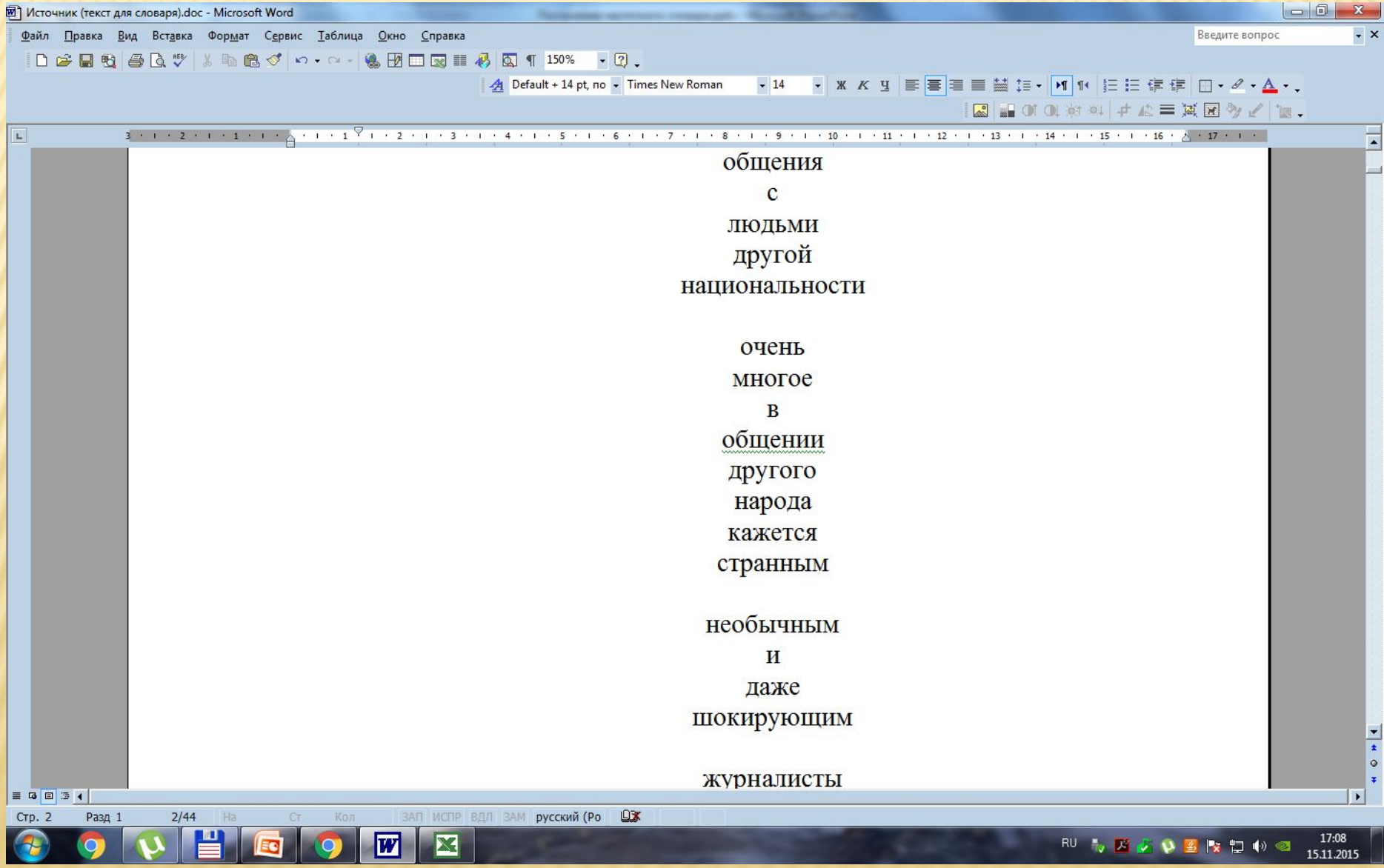
Рис. 4


Источник (текст для словаря).doc - Microsoft Word
Файл Правка Вид Вставка Формат Сервис Таблица Окно Справка
Введите вопрос
принято
разговаривать
с
попутчиками
у
ру
не
Не
в
ли
зд
и
прощаются
со
своими
попутчиками
Найти и заменить
Найти
Заменить
Перейти
Найти:
Заменить на:
Microsoft Word
Поиск документа закончен. Произведено замен: 1 916.
OK
Параметры поиска
Направление:
Везде
Учитывать регистр
Только слово целиком
Подстановочные знаки
Произносится как
Все словоформы
Учитывать диакр. знаки
Учитывать алиф гамза
Учитывать непечатаемые знаки
Произносится как (японский)
Параметры...
Заменить
Формат
Специальный
Снять форматирование
Замена:
Нажмите Esc для отмены.
RU
16:09
15.11.2015

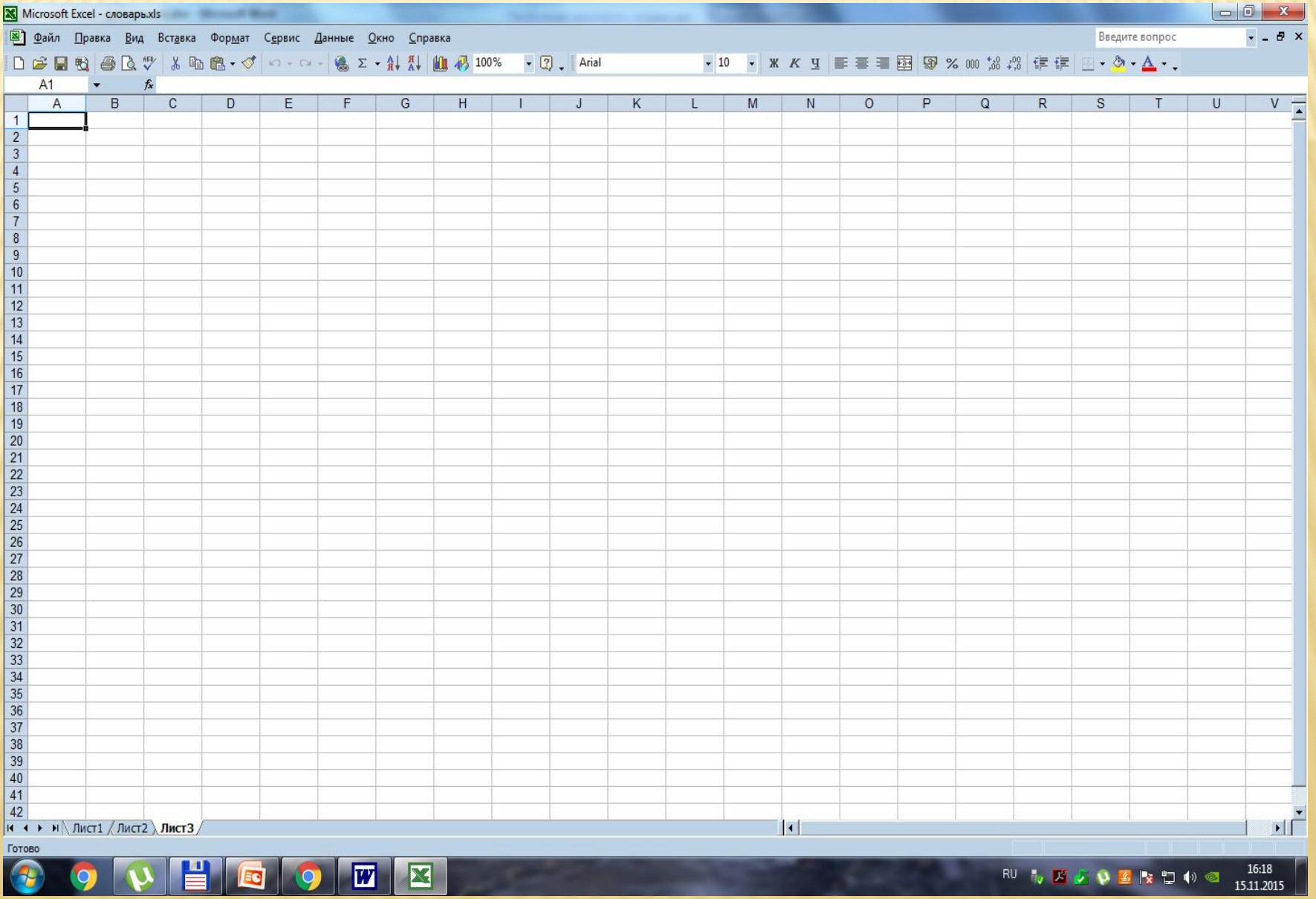
3. Далее сориентируем все слова по центру листа (Рис. 5).



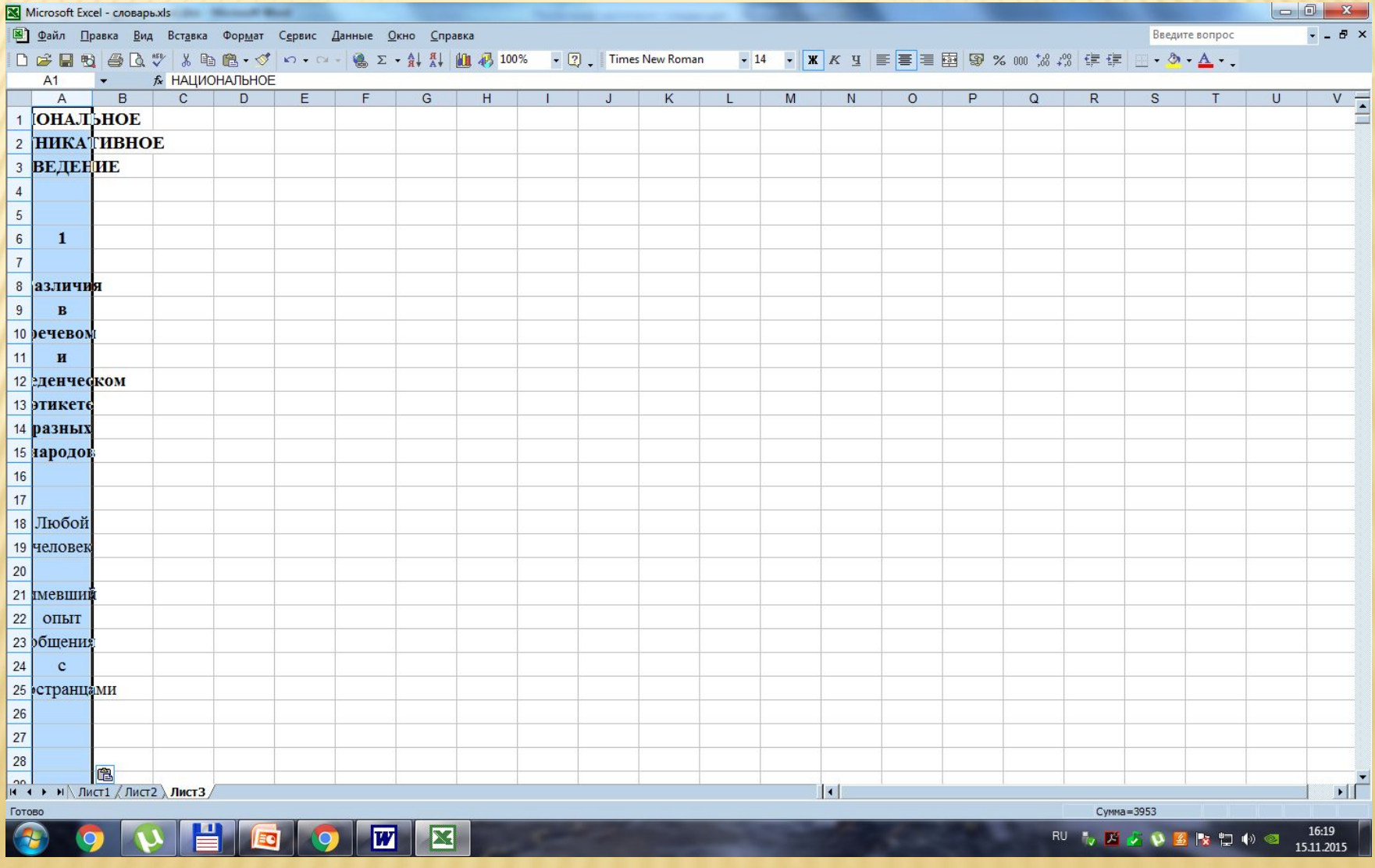
4. С помощью сочетания клавиш Shift и F3 уберем прописные буквы (Рис. 6).



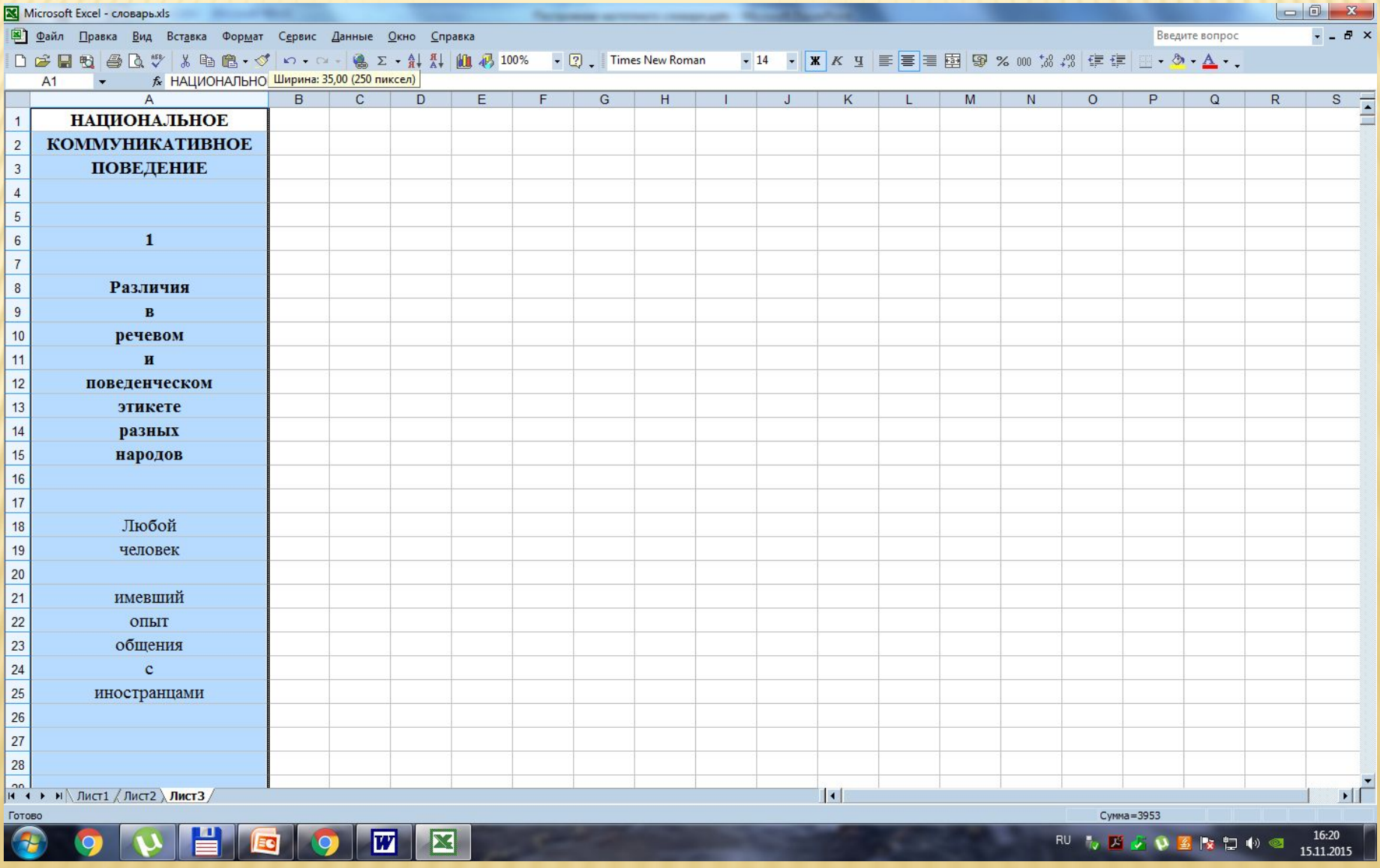
5. Создаем документ Microsoft Excel (Рис. 7).



6. Вставляем скопированные слова в таблицу Microsoft Excel (Рис. 8).

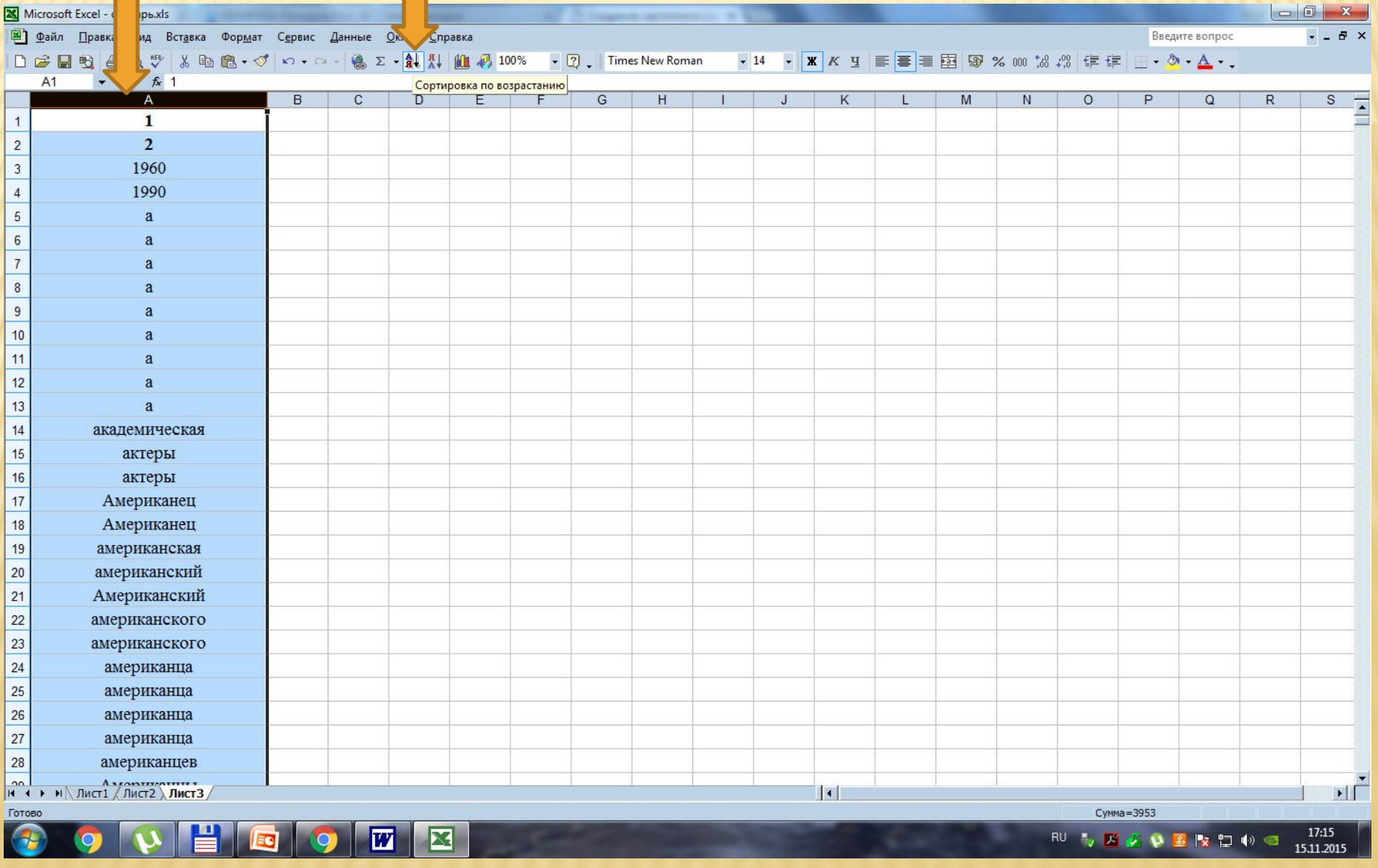


7. Создаем необходимую ширину столбца (Рис. 9).

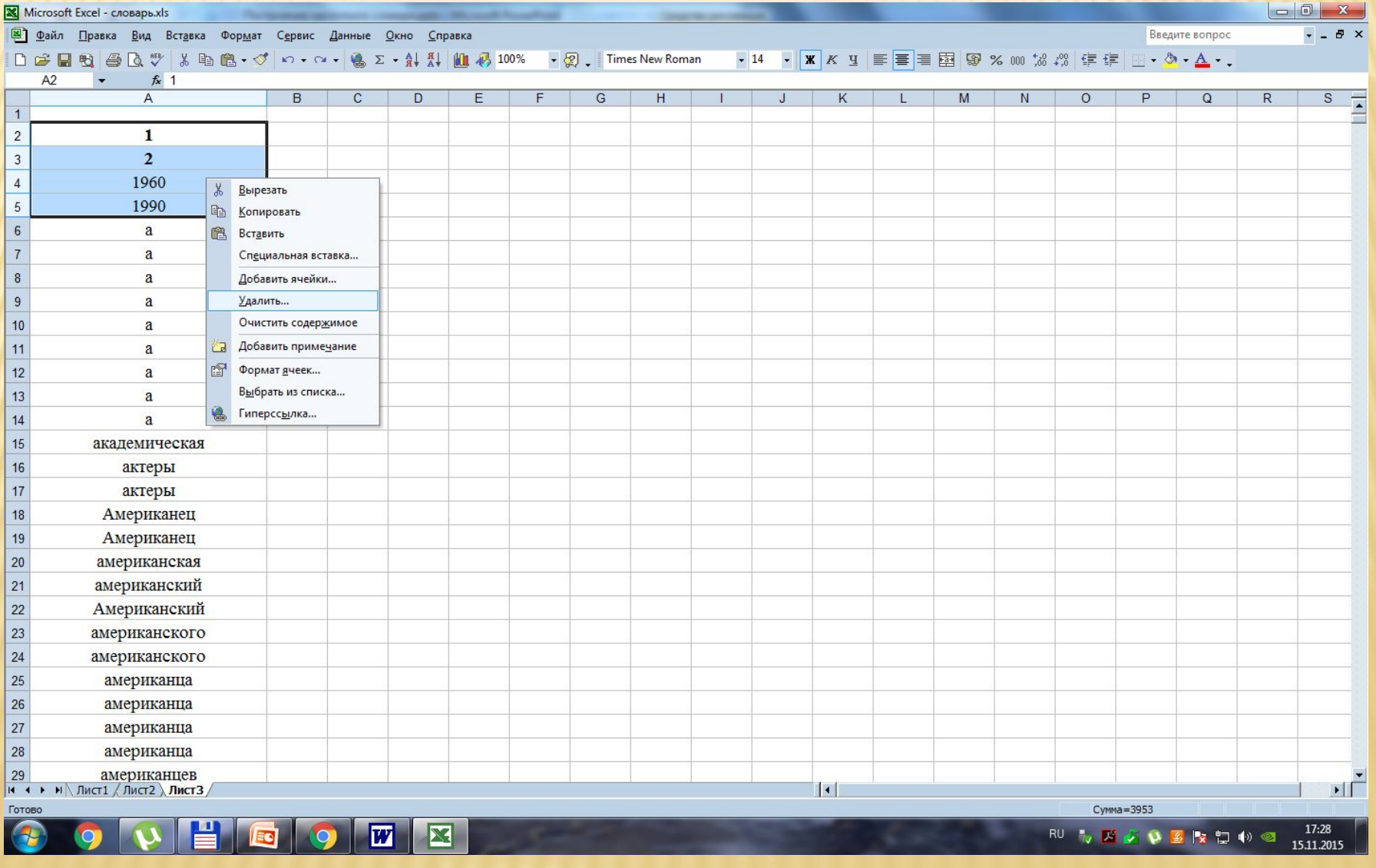


8. Далее сортируем список слов в таблице по алфавиту. Для этого необходимо выделить их нажатием на колонку «А». Затем использовать опцию сортировка по возрастанию (от А до Я). (Рис. 10)

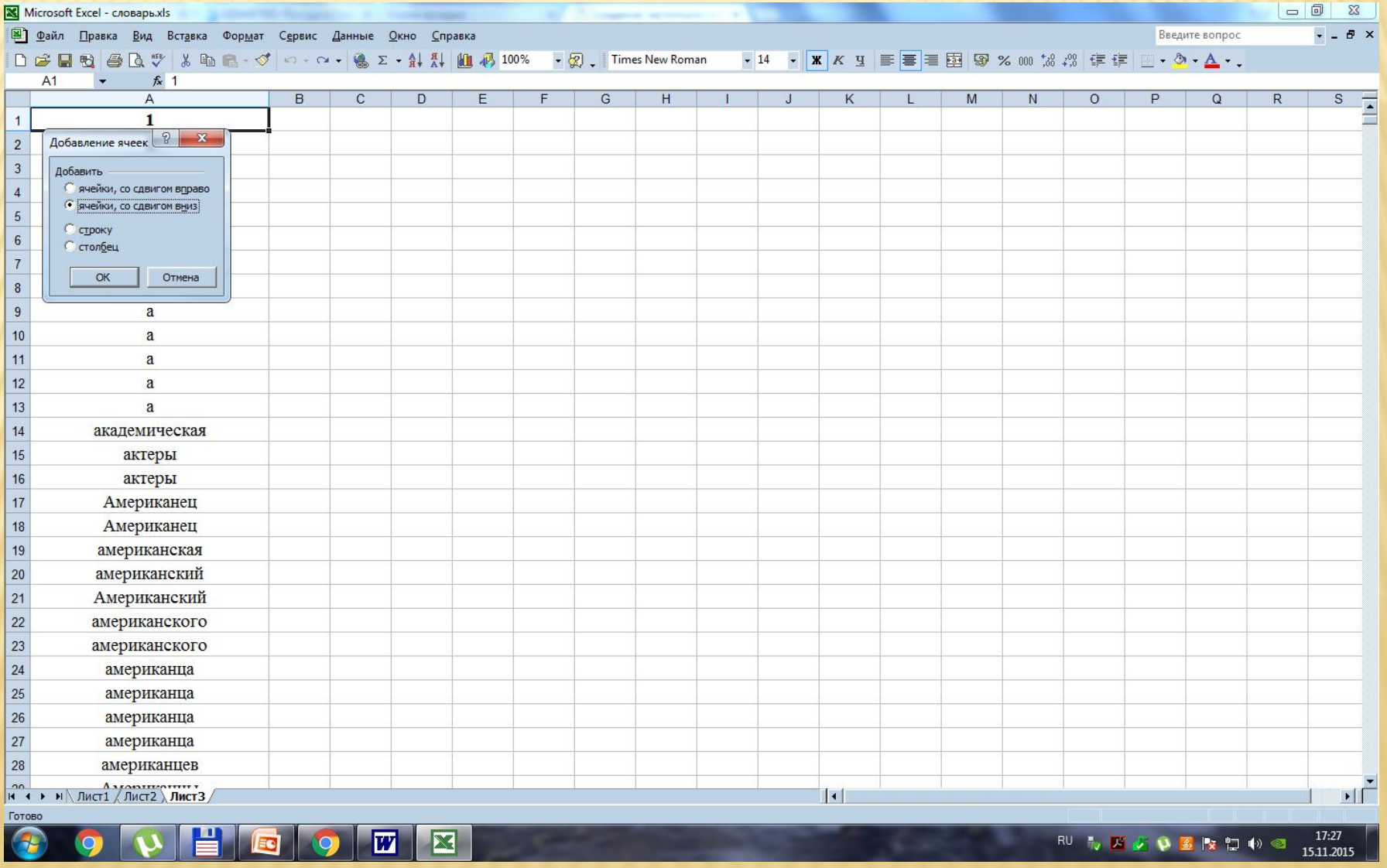
Рис. 10


Microsoft Excel
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
Введите вопрос
Times New Roman
14
100%
Сортировка по возрастанию
A1
1
1
2
1960
1990
a
a
a
a
a
a
a
a
a
академическая
актеры
актеры
Американец
Американец
американская
американский
Американский
американского
американского
американца
американца
американца
американца
американцев
Лист1 Лист2 Лист3
Готово
Сумма=3953
RU
17:15
15.11.2015

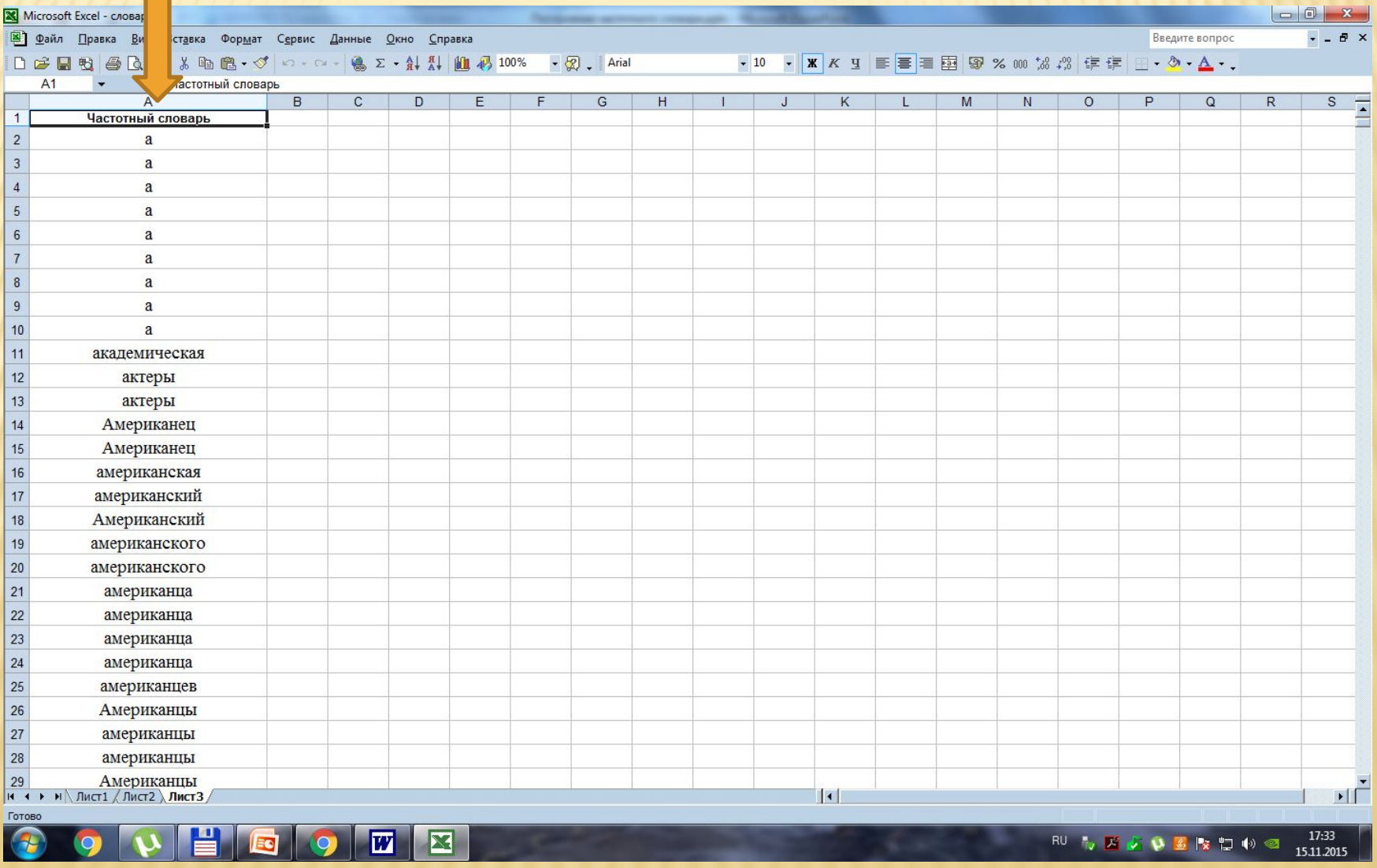
9. При необходимости можно удалить ячейки с цифровыми данными (Рис. 11).



10. Затем нажатием правой кнопки мыши по первой ячейке добавляем новую ячейку (Рис. 12).

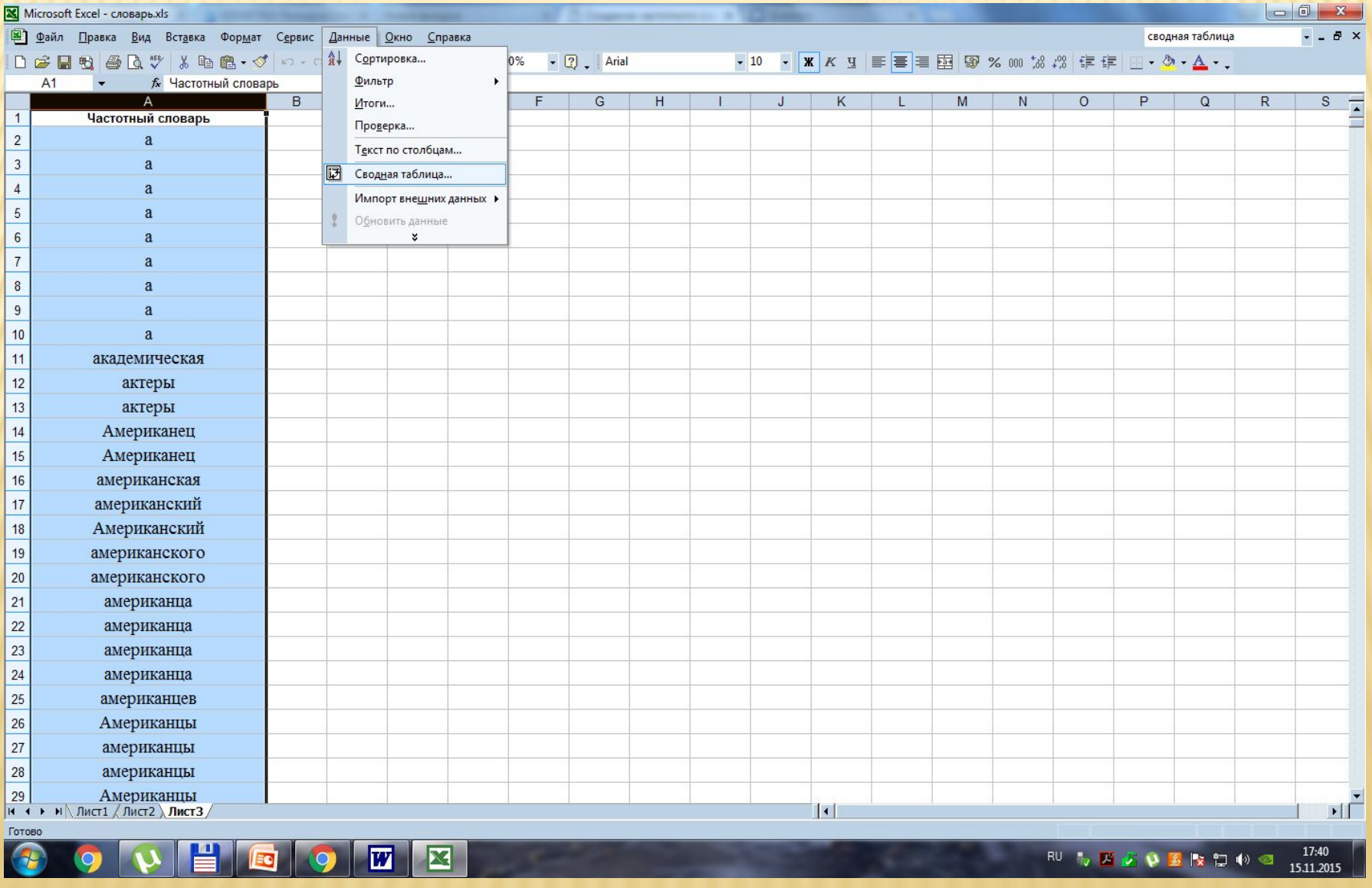


11. Затем называем нашу таблицу. Например, «Частотный словарь» (Рис. 13).



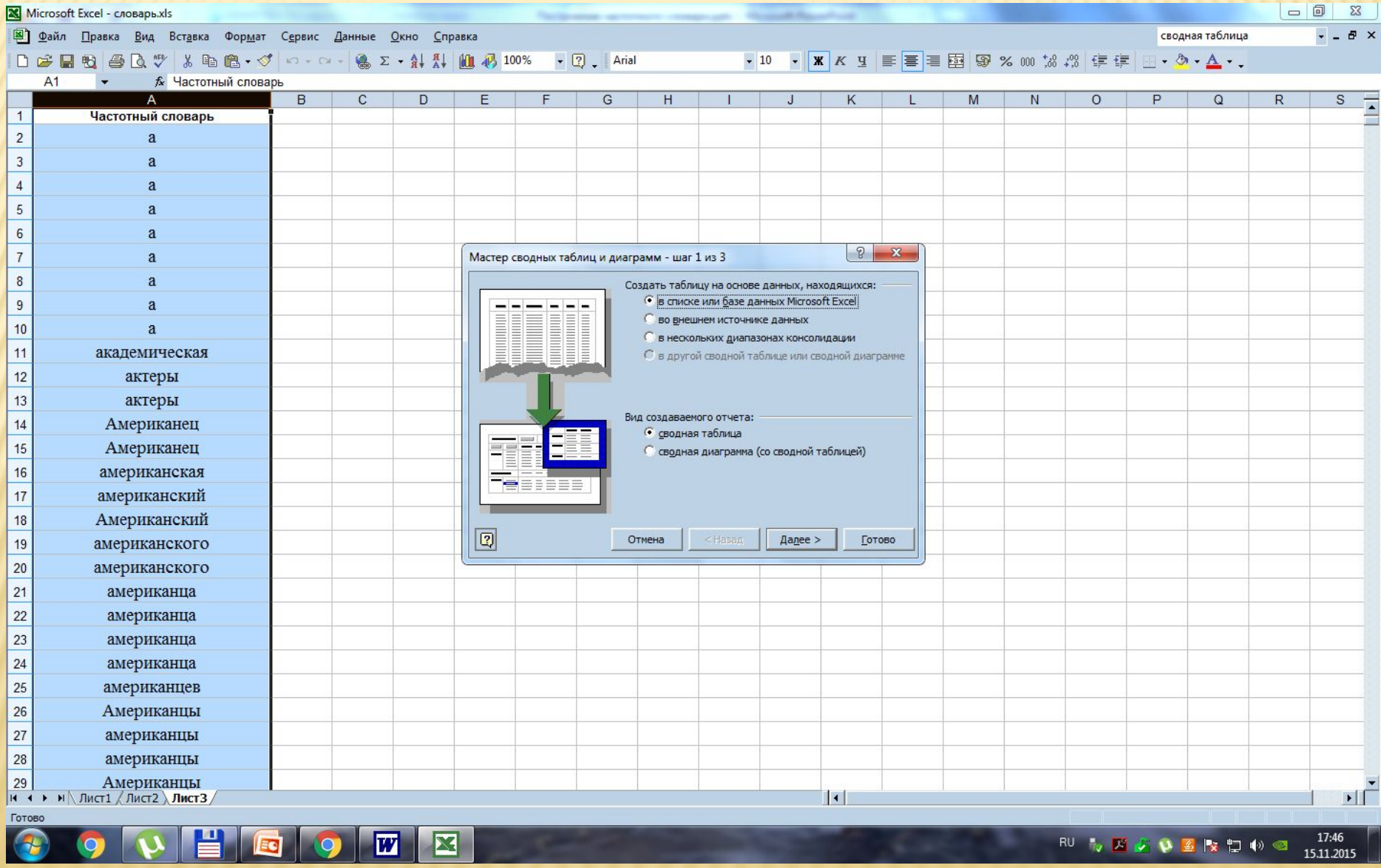
12. Далее выделяем столбец нажатием на колонку «А» и создаем сводную таблицу на новом листе. Для этого переходим на вкладку «Данные» → «Сводная таблица». (Рис. 14)

Рис. 14.


Microsoft Excel - словарь.xls
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
сводная таблица
Сортировка...
Фильтр
Итоги...
Проверка...
Текст по столбцам...
Сводная таблица...
Импорт внешних данных
Обновить данные
Arial
10
A1
Частотный словарь
Частотный словарь
а
академическая
актеры
актеры
Американец
Американец
американская
американский
Американский
американского
американского
американца
американца
американца
американца
американцев
Американцы
американцы
американцы
Американцы
Лист1 Лист2 Лист3
Готово
RU
17:40
15.11.2015

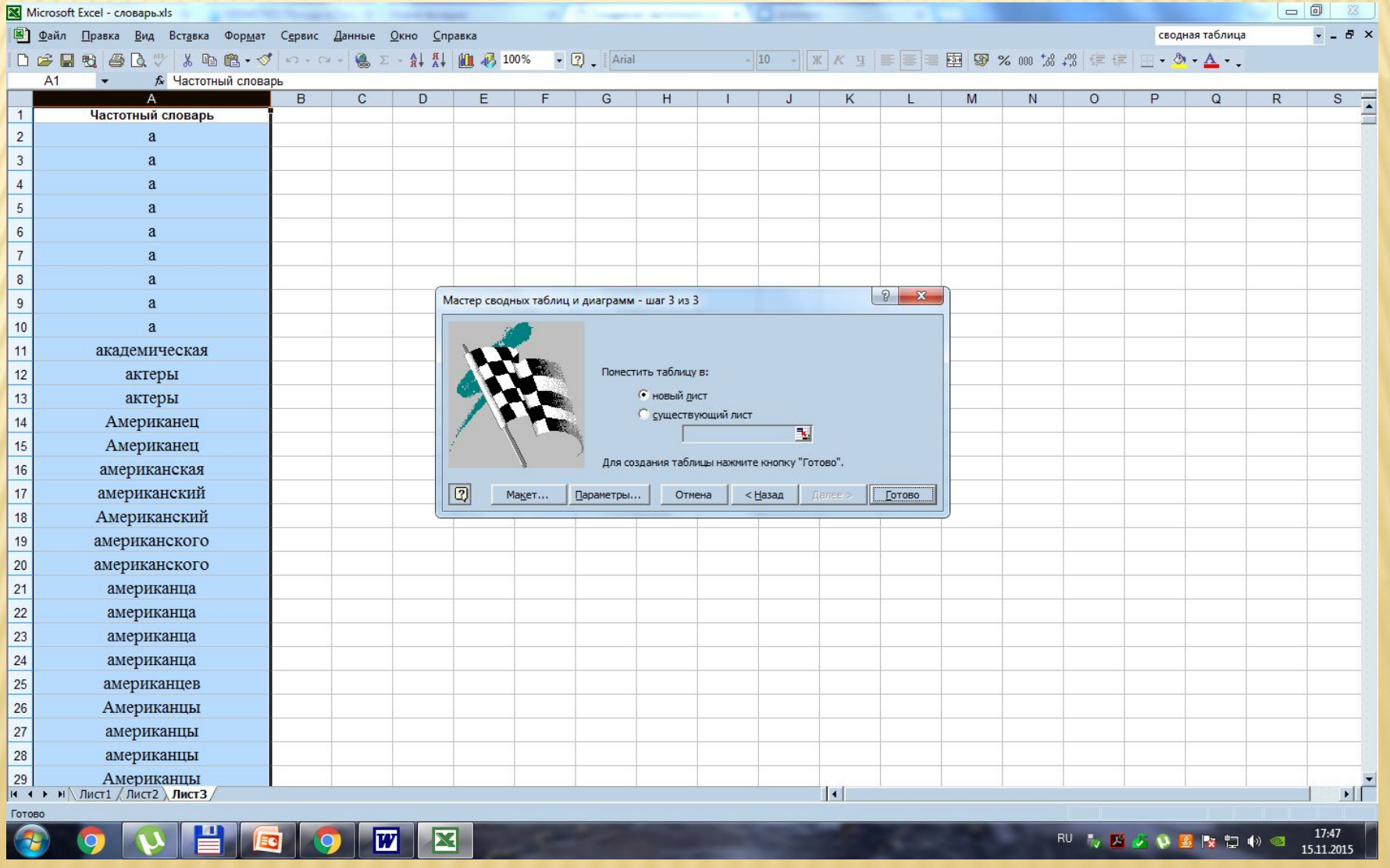
13. Затем соглашаемся с предлагаемыми в диалоговом окне параметрами и заканчиваем создание сводной таблицы на новом листе (Рис. 15-17).

Рис. 15



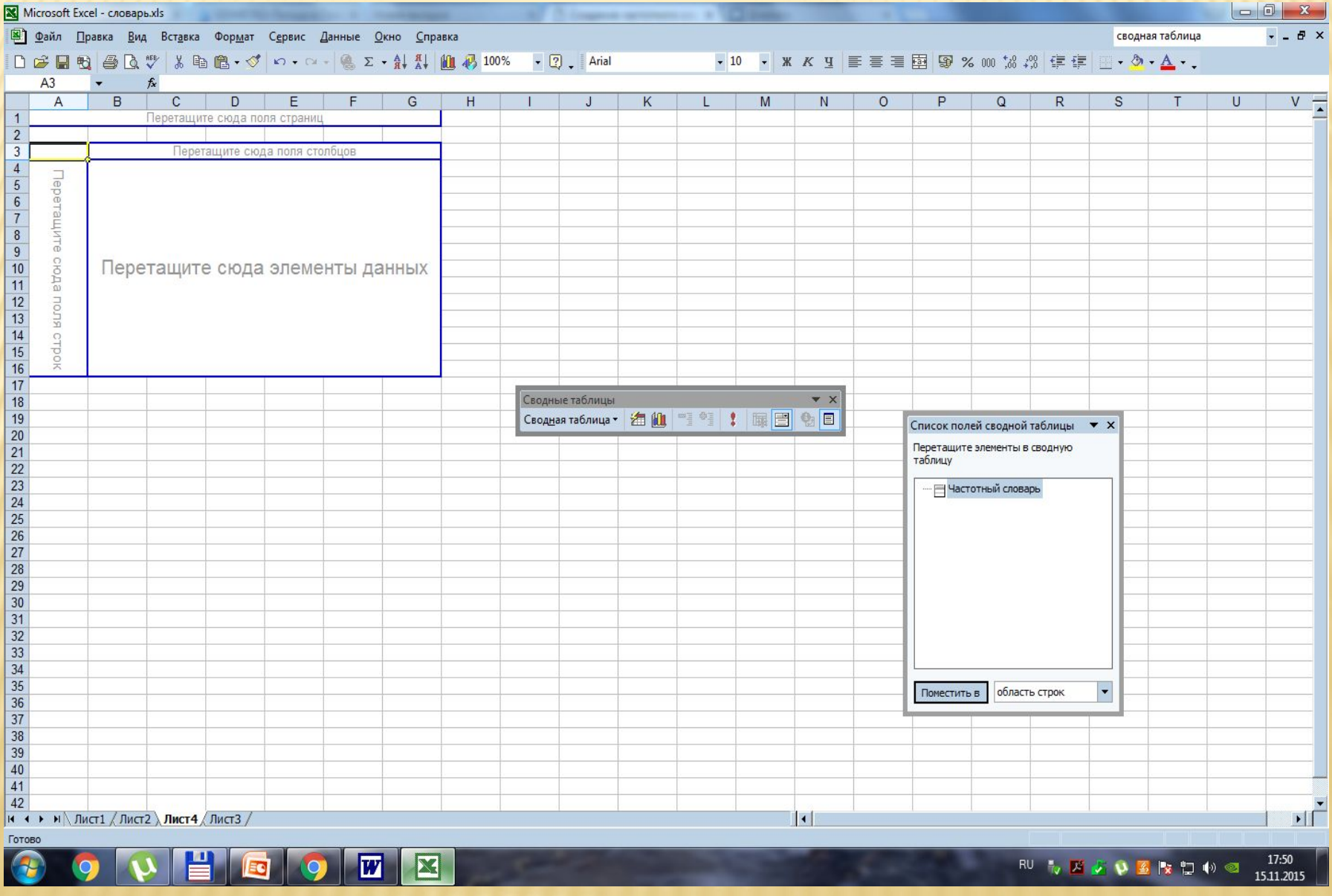
	A
1	Частотный словарь
2	а
3	а
4	а
5	а
6	а
7	а
8	а
9	а
10	а
11	академическая
12	актеры
13	актеры
14	Американец
15	Американец
16	американская
17	американский
18	Американский
19	американского
20	американского
21	американца
22	американца
23	американца
24	американца
25	американцев
26	Американцы
27	американцы
28	американцы
29	Американцы

Рис. 16.



	A
1	Частотный словарь
2	а
3	а
4	а
5	а
6	а
7	а
8	а
9	а
10	а
11	академическая
12	актеры
13	актеры
14	Американец
15	Американец
16	американская
17	американский
18	Американский
19	американского
20	американского
21	американца
22	американца
23	американца
24	американца
25	американцев
26	Американцы
27	американцы
28	американцы
29	Американцы

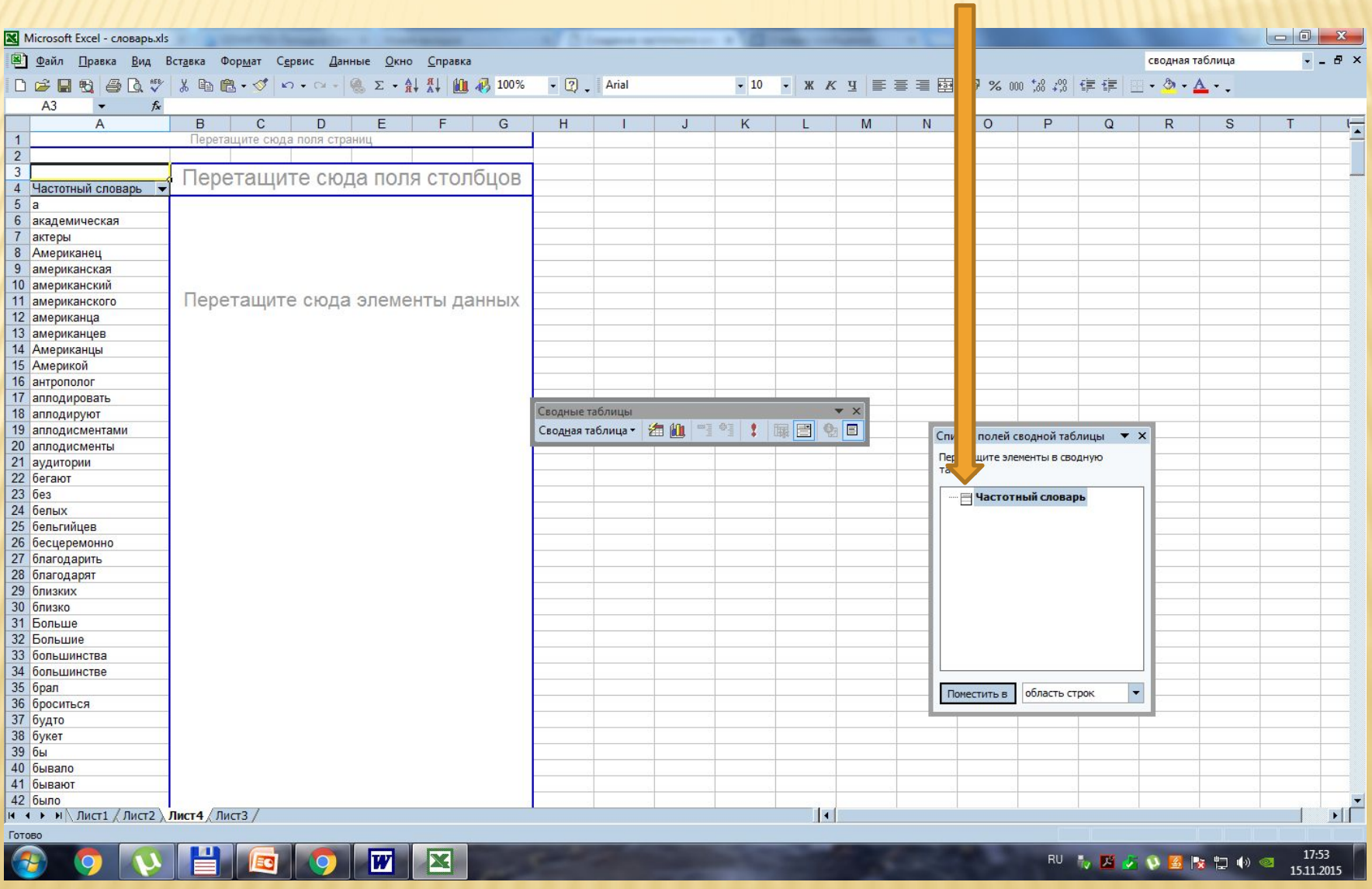
Рис. 17.


Microsoft Excel - словарь.xls
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
сводная таблица
Arial
10
100%
A3
Перетащите сюда поля страниц
Перетащите сюда поля столбцов
Перетащите сюда поля строк
Перетащите сюда элементы данных
Сводные таблицы
Сводная таблица
Список полей сводной таблицы
Перетащите элементы в сводную таблицу
Частотный словарь
Поместить в
область строк
Лист1 Лист2 Лист4 Лист3
Готово
RU
17:50
15.11.2015

14. После этого дважды щелкаем левой кнопкой мыши по значку «Частотный словарь» и в левой части листа появляется сводная таблица (Рис. 18).

В отличие от предыдущего списка слов, в сводной таблице каждая словоформа встречается только один раз.

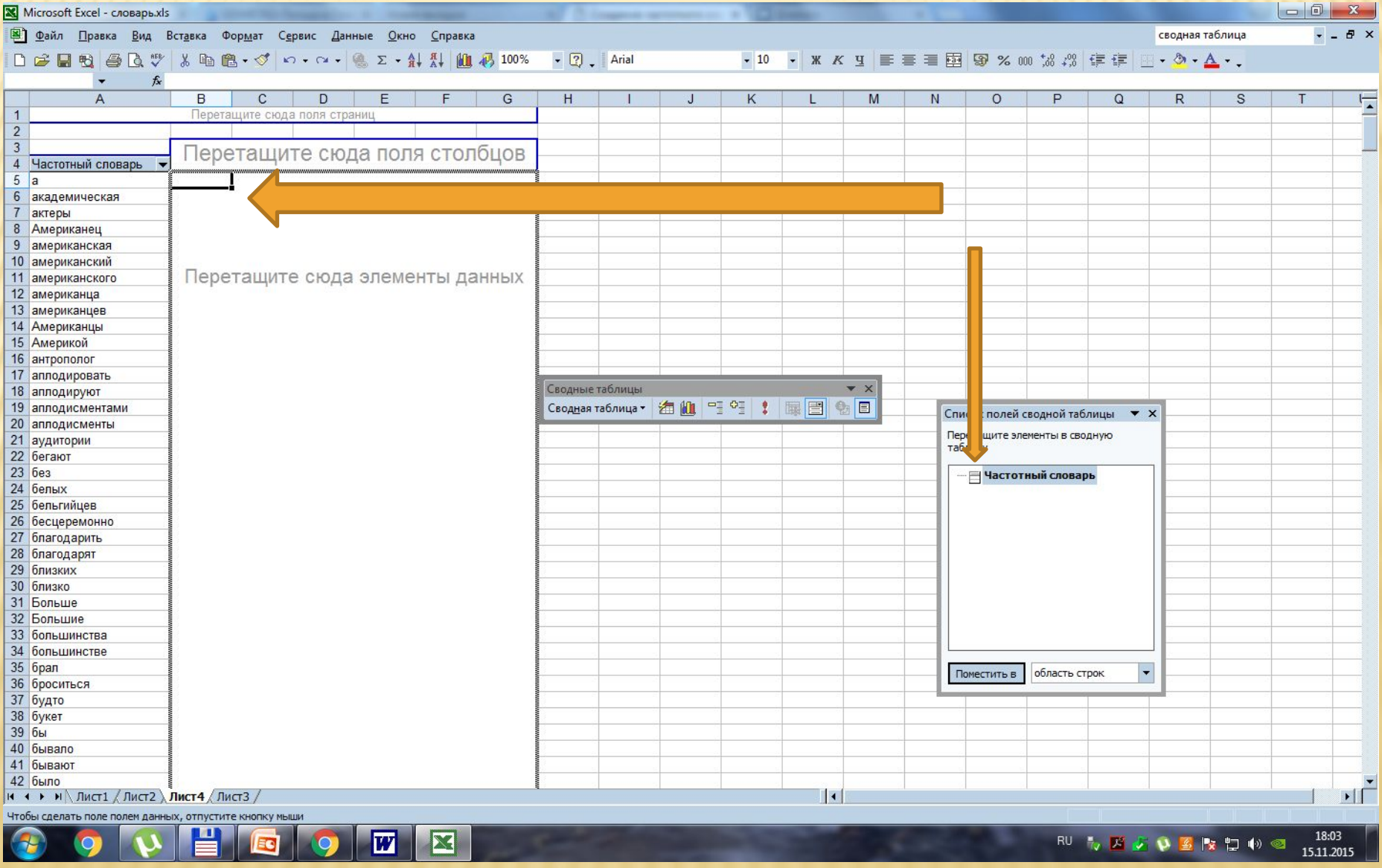
Рис. 18.


Microsoft Excel - словарь.xls
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
сводная таблица
Перетащите сюда поля страниц
Перетащите сюда поля столбцов
Частотный словарь
Перетащите сюда элементы данных
Сводные таблицы
Сводная таблица
Частотный словарь
Поместить в
область строк
Готово
RU
17:53
15.11.2015

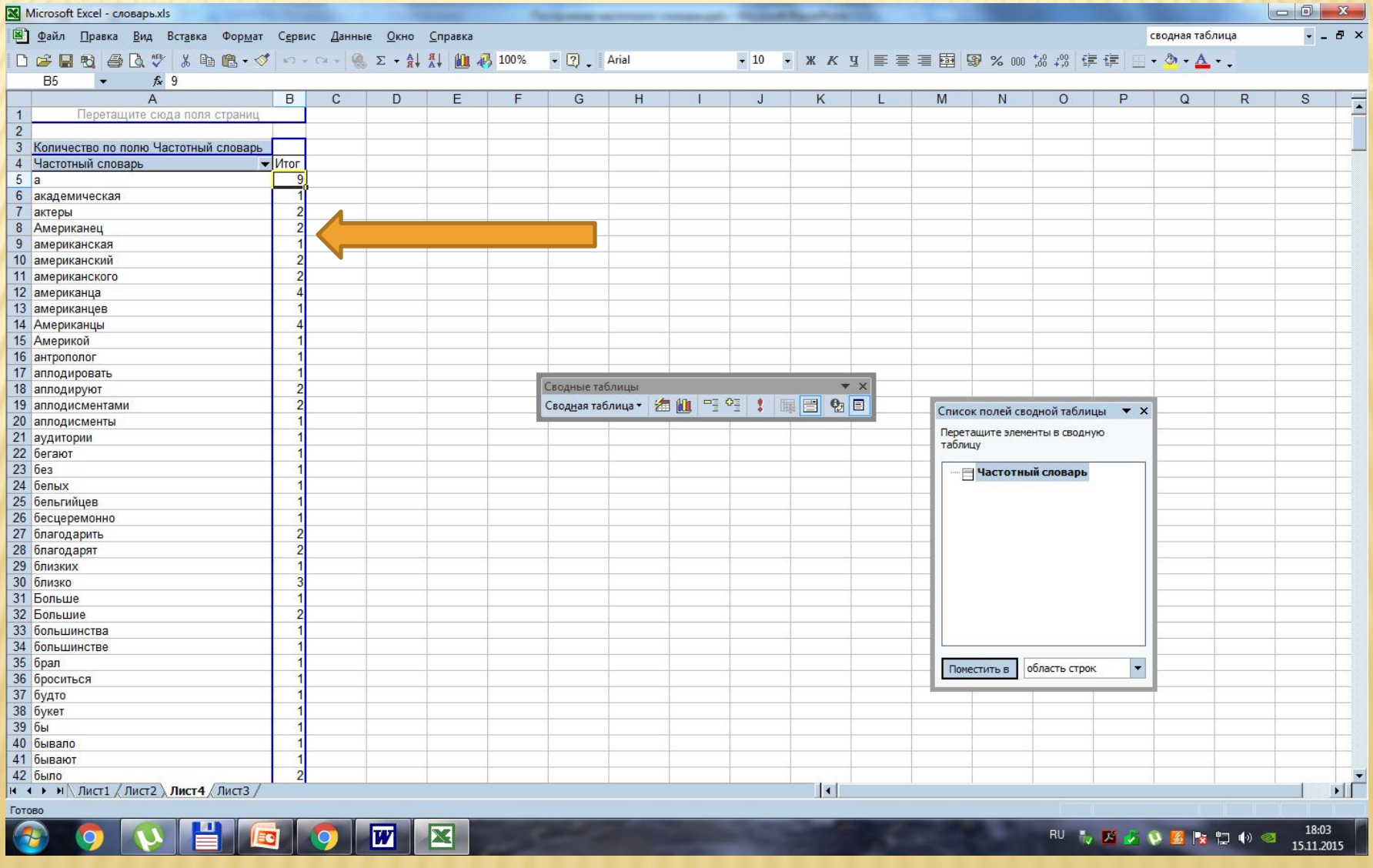
15. Для получения полноценного частотного словаря в полученной сводной таблице необходимо отобразить столбец с частотой слов.

Для этого перетаскиваем значок «Частотный словарь», нажав на него левой кнопки мыши, в пустую выделенную область справа от списка слов (Рис. 19).

Рис. 19

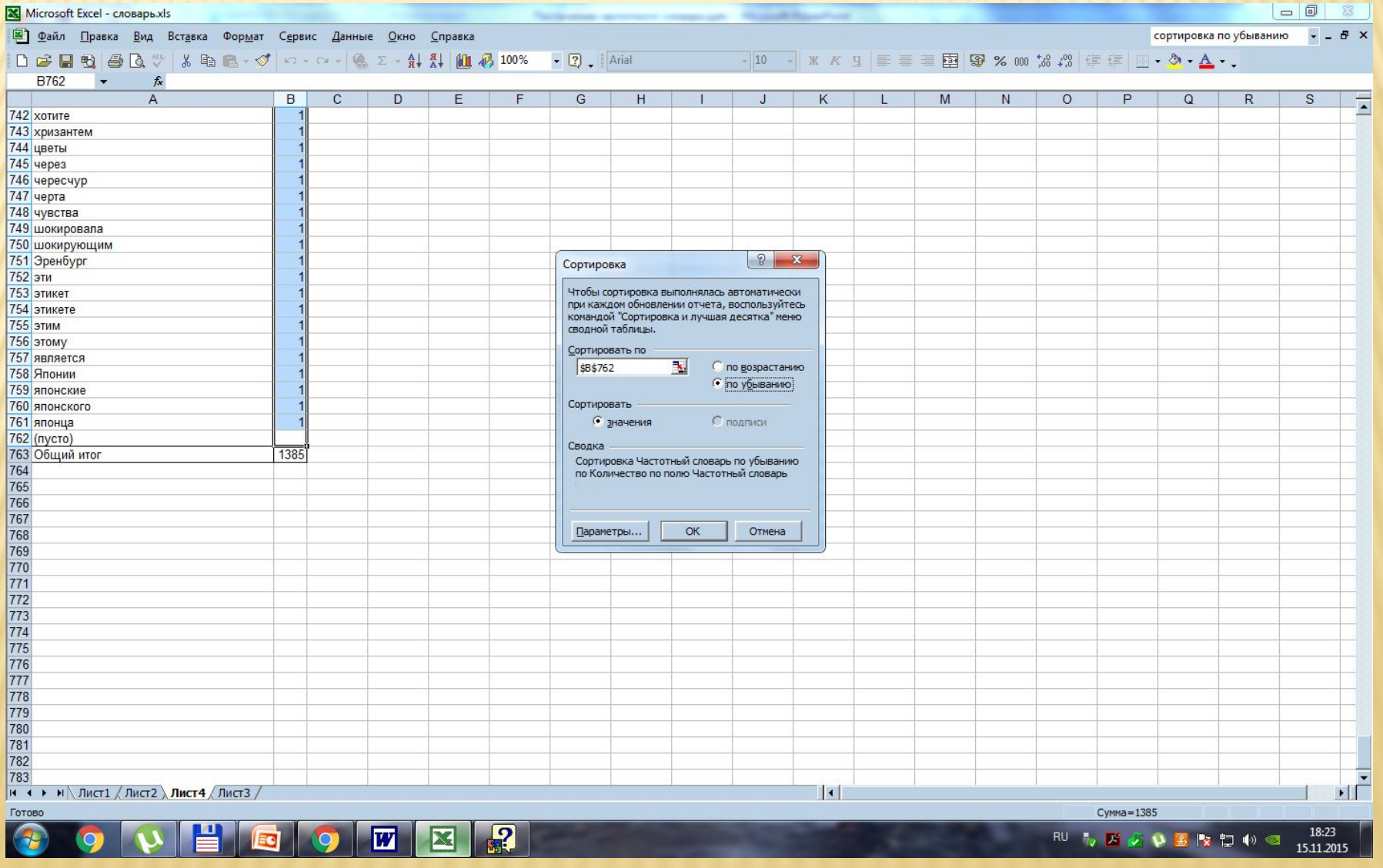
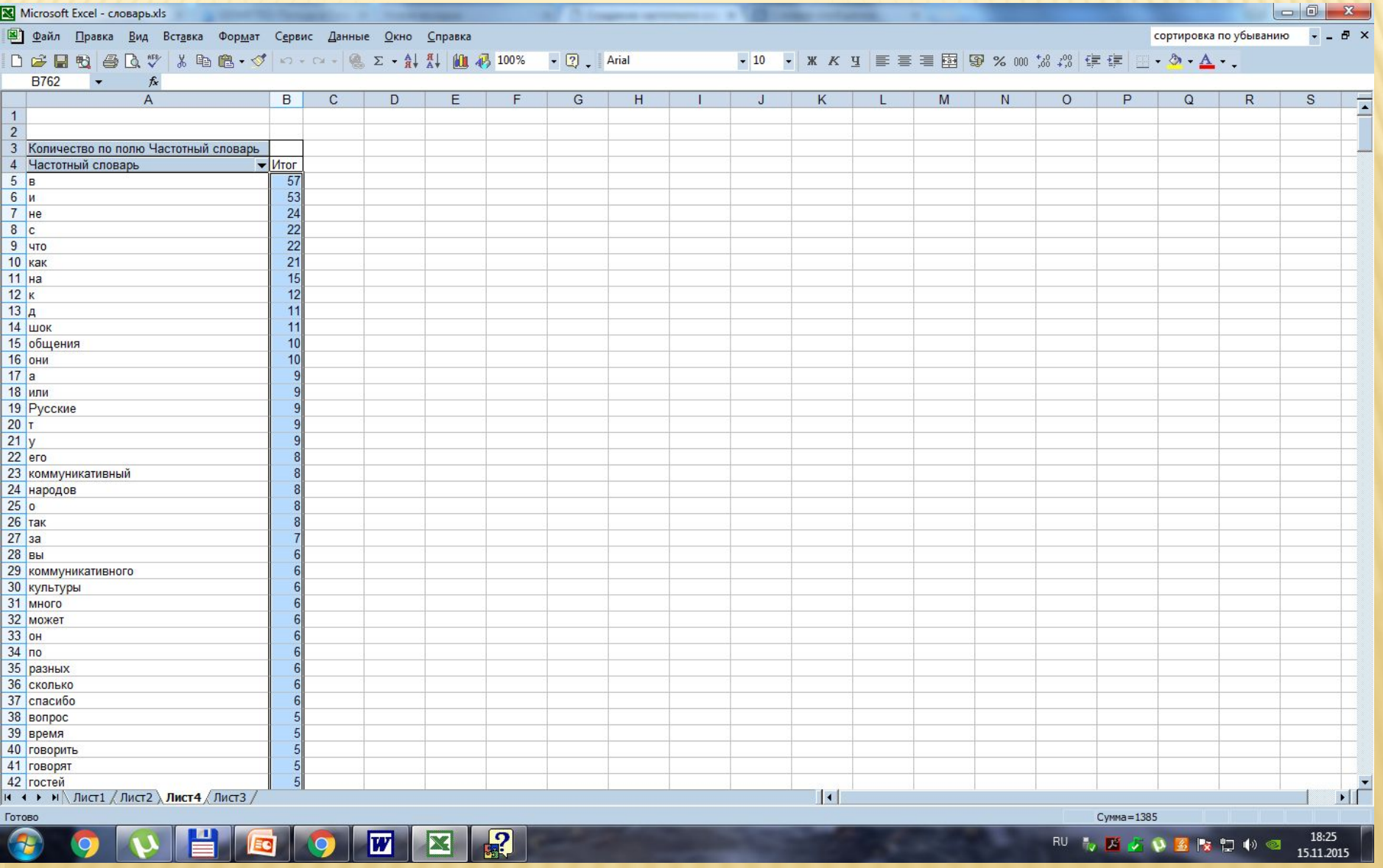

Microsoft Excel - словарь.xls
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
сводная таблица
Перетащите сюда поля страниц
Перетащите сюда поля столбцов
Частотный словарь
Перетащите сюда элементы данных
Сводные таблицы
Сводная таблица
Частотный словарь
Поместить в
область строк
Лист1 Лист2 Лист4 Лист3
Чтобы сделать поле полем данных, отпустите кнопку мыши
RU
18:03
15.11.2015

16. После этого отобразиться столбец с частотой слов (Рис. 20).



17. Далее выделяем столбец с числовыми значениями (лучше выделять от последнего слова к первому). Затем копируем выделение. После этого нажимаем на вкладку «Данные» «Сортировка». Выбираем позицию «По убыванию» (Рис. 21).

Рис. 21


Microsoft Excel - словарь.xls
Файл Правка Вид Вставка Формат Сервис Данные Окно Справка
сортировка по убыванию
B762
742 хотите 1
743 хризантем 1
744 цветы 1
745 через 1
746 чересчур 1
747 черта 1
748 чувства 1
749 шокировала 1
750 шокирующим 1
751 Эренбург 1
752 эти 1
753 этикет 1
754 этикете 1
755 этим 1
756 этому 1
757 является 1
758 Японии 1
759 японские 1
760 японского 1
761 японца 1
762 (пусто)
763 Общий итог 1385
Сортировка
Чтобы сортировка выполнялась автоматически при каждом обновлении отчета, воспользуйтесь командой "Сортировка и лучшая десятка" меню сводной таблицы.
Сортировать по
B762
по возрастанию
по убыванию
Сортировать
значения
подписи
Сводка
Сортировка Частотный словарь по убыванию по Количество по полю Частотный словарь
Параметры...
OK
Отмена
Лист1 Лист2 Лист4 Лист3
Готово
Сумма=1385
RU
18:23
15.11.2015

Вы получили частотный словарь анализируемого текста.