

**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

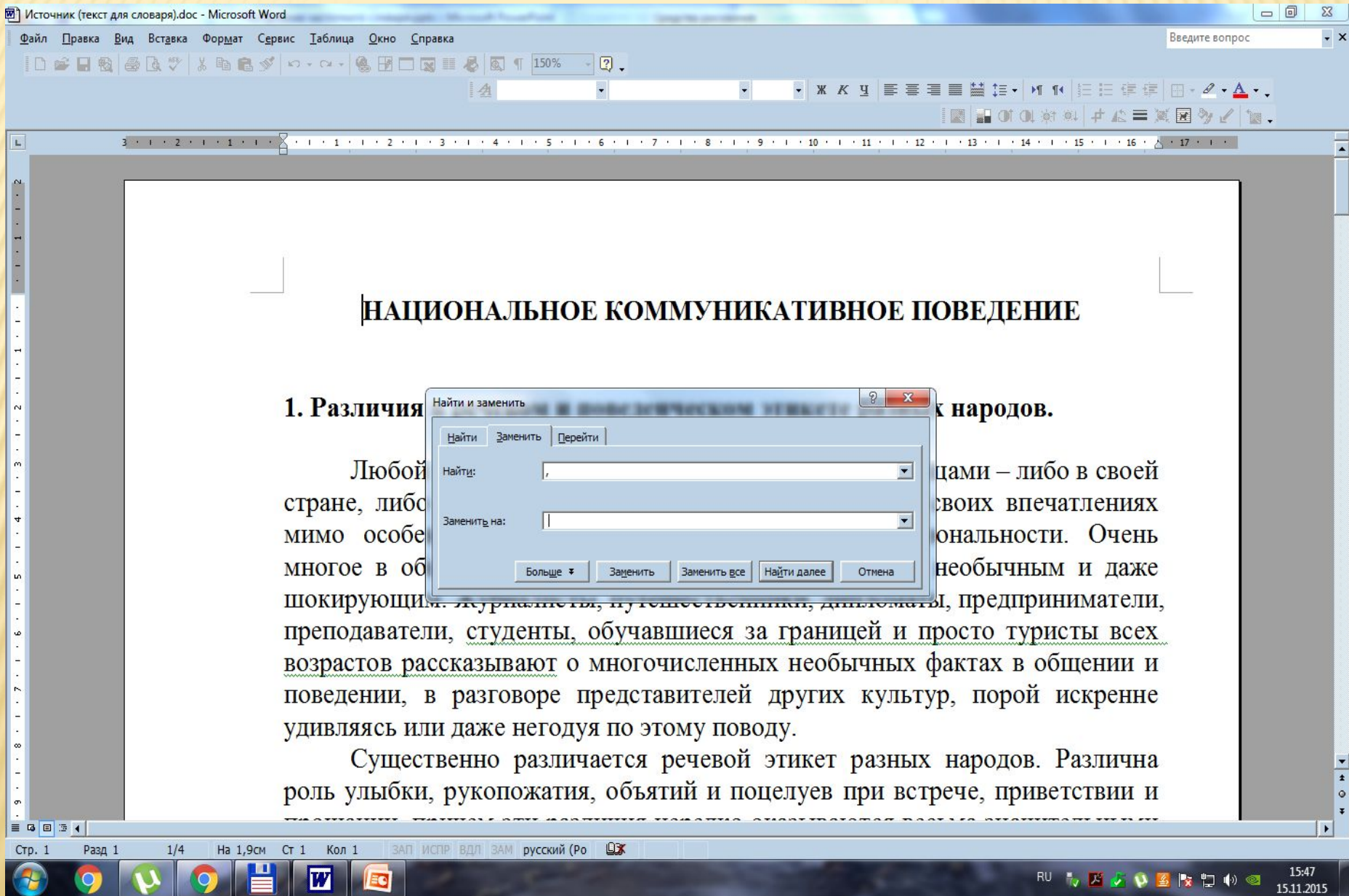
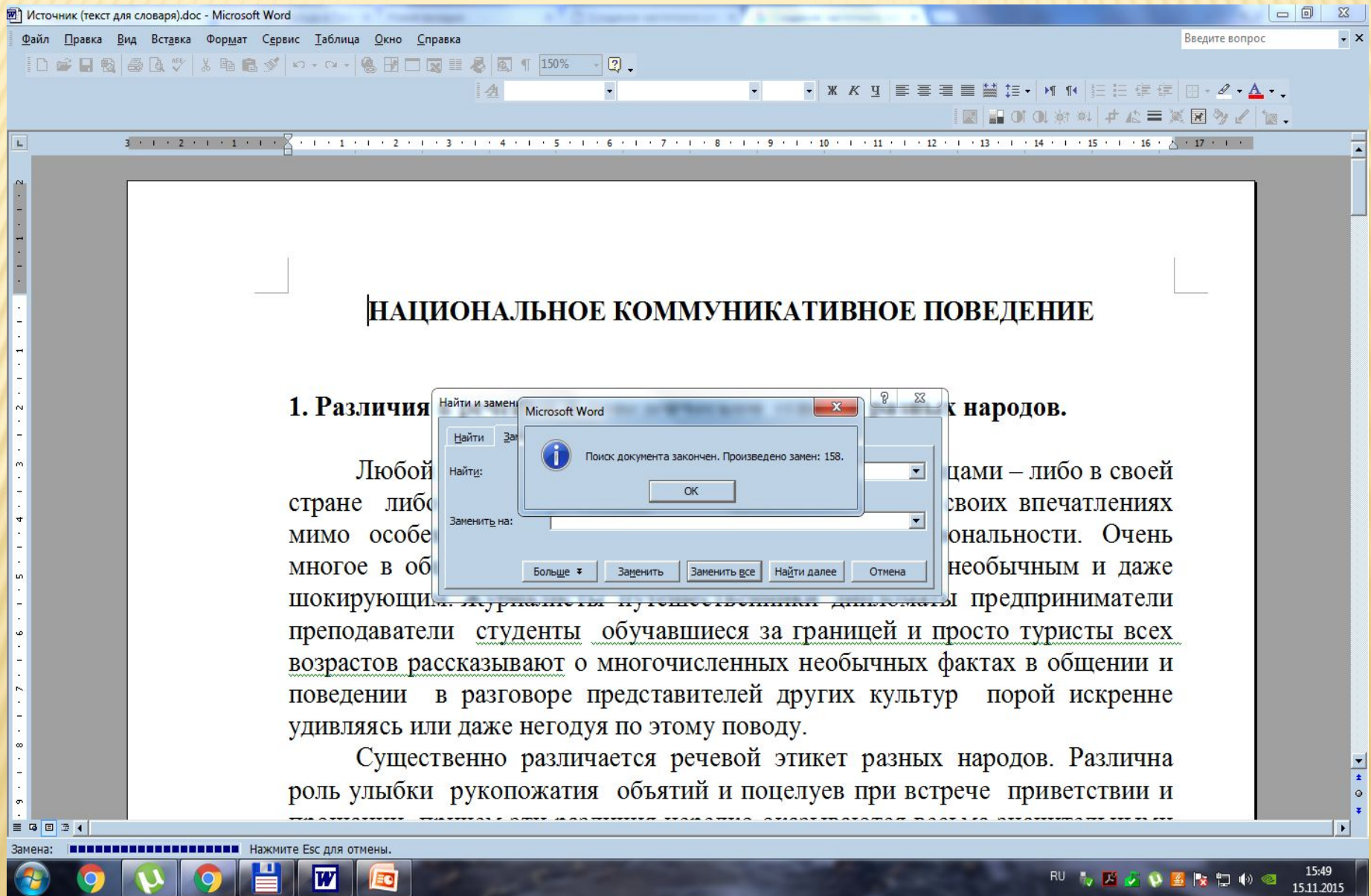


Рис. 2.



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

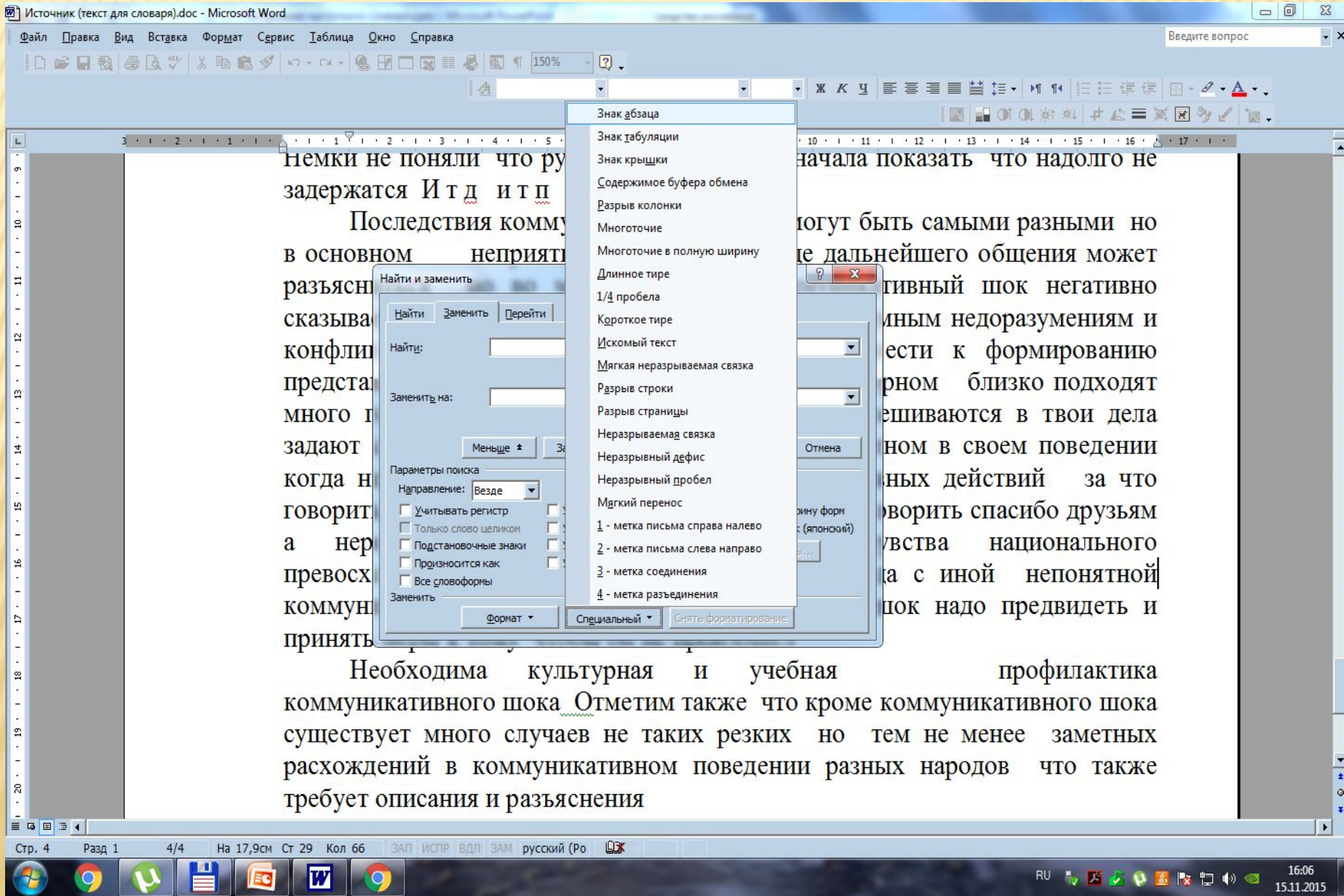
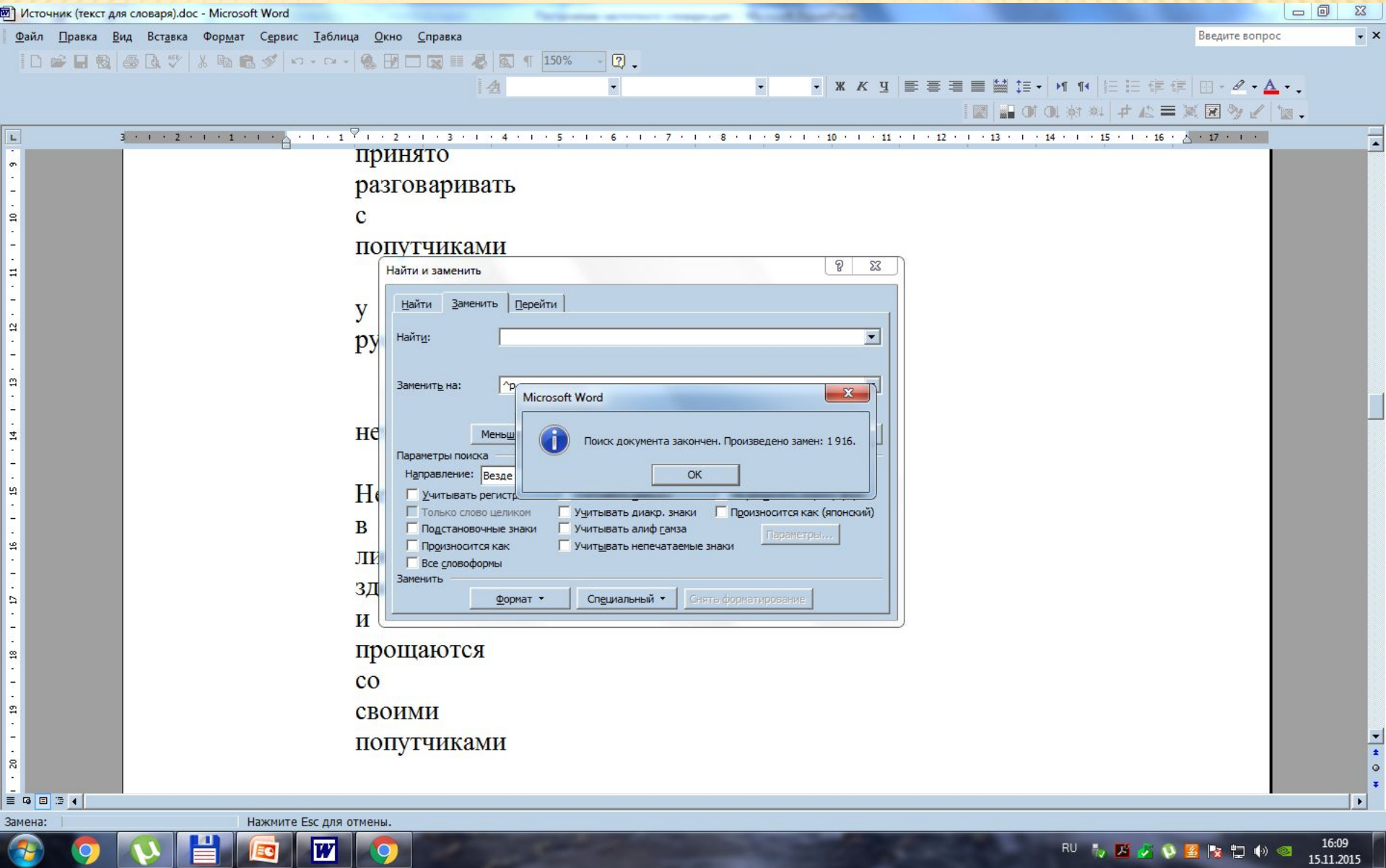
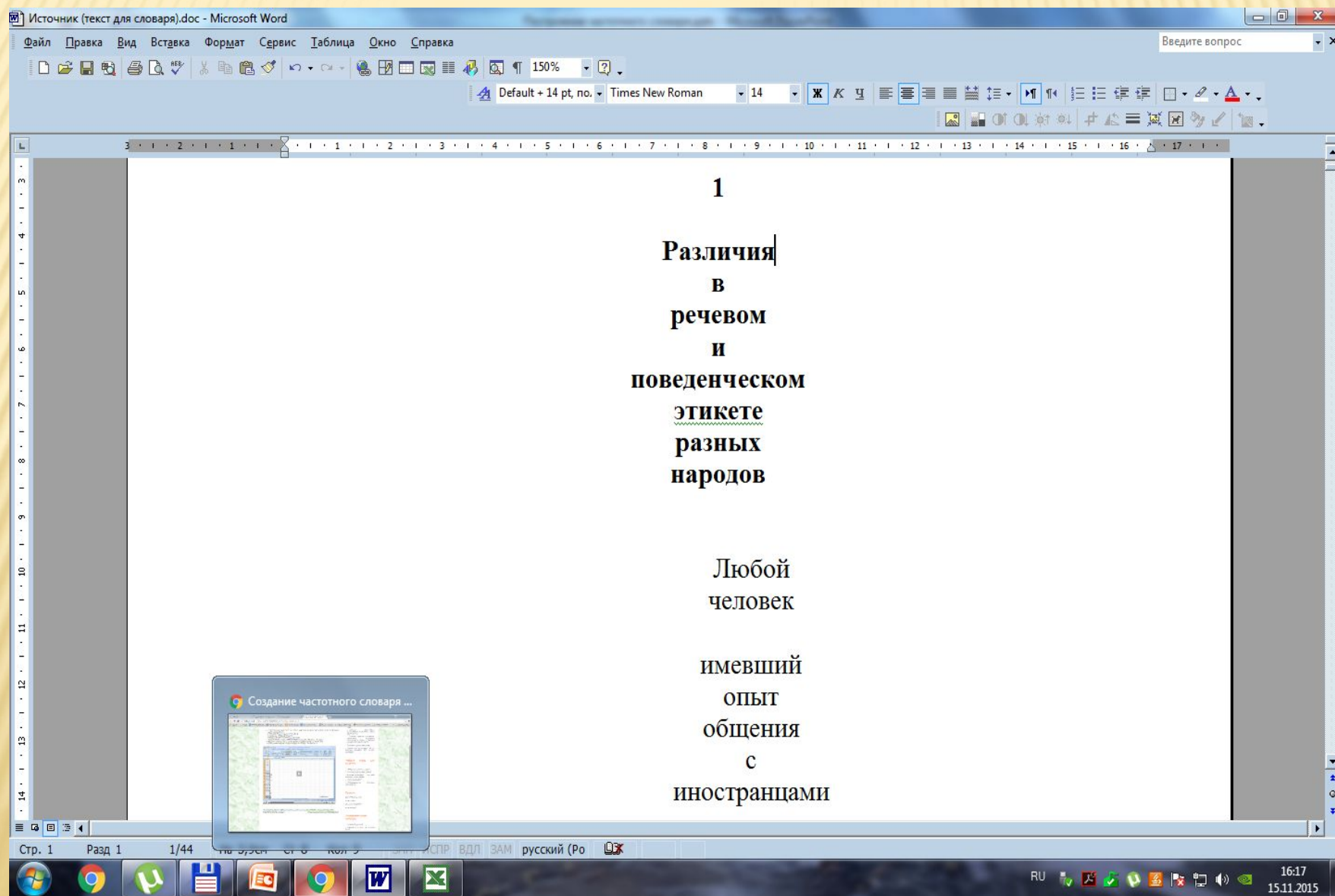


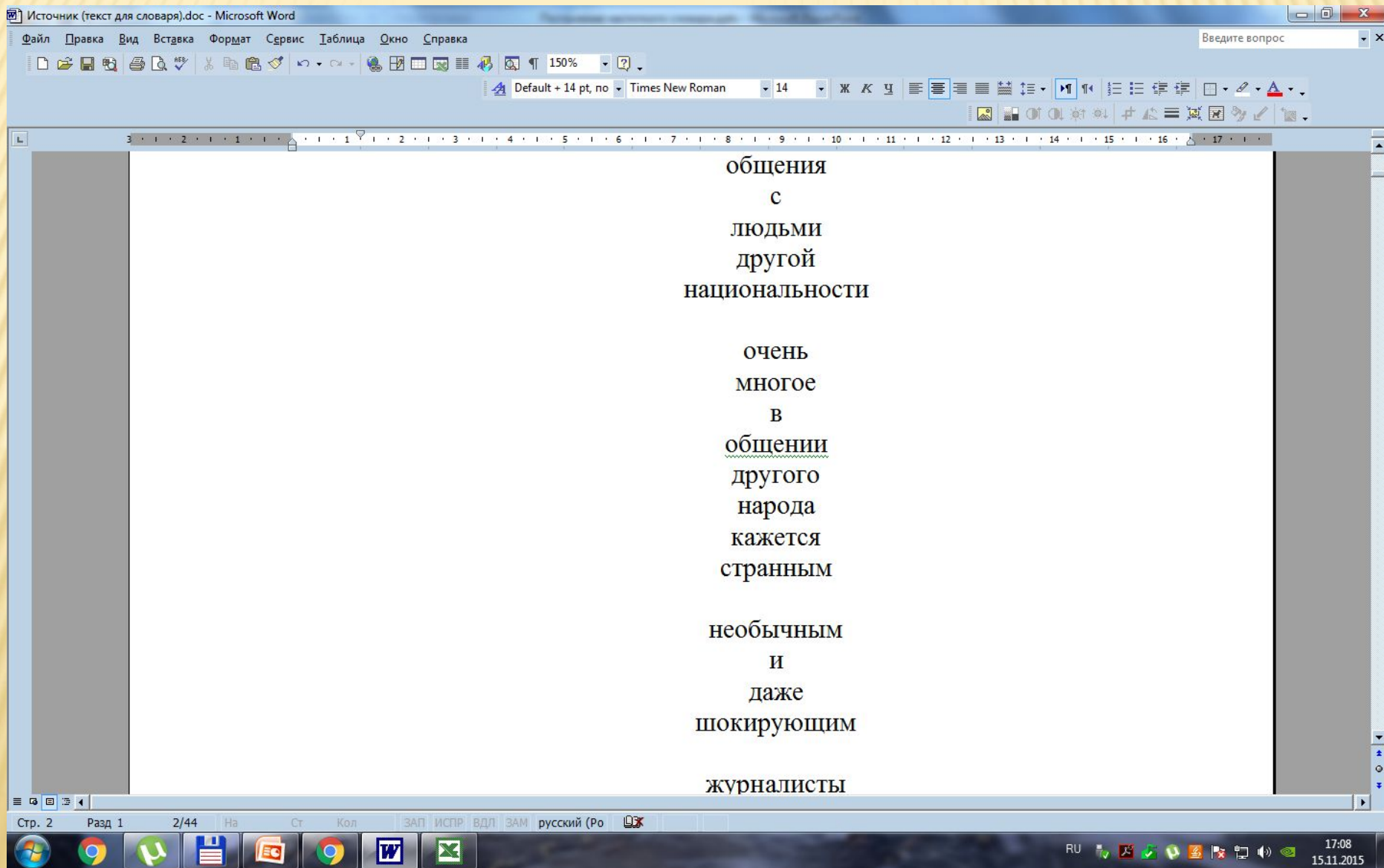
Рис. 4



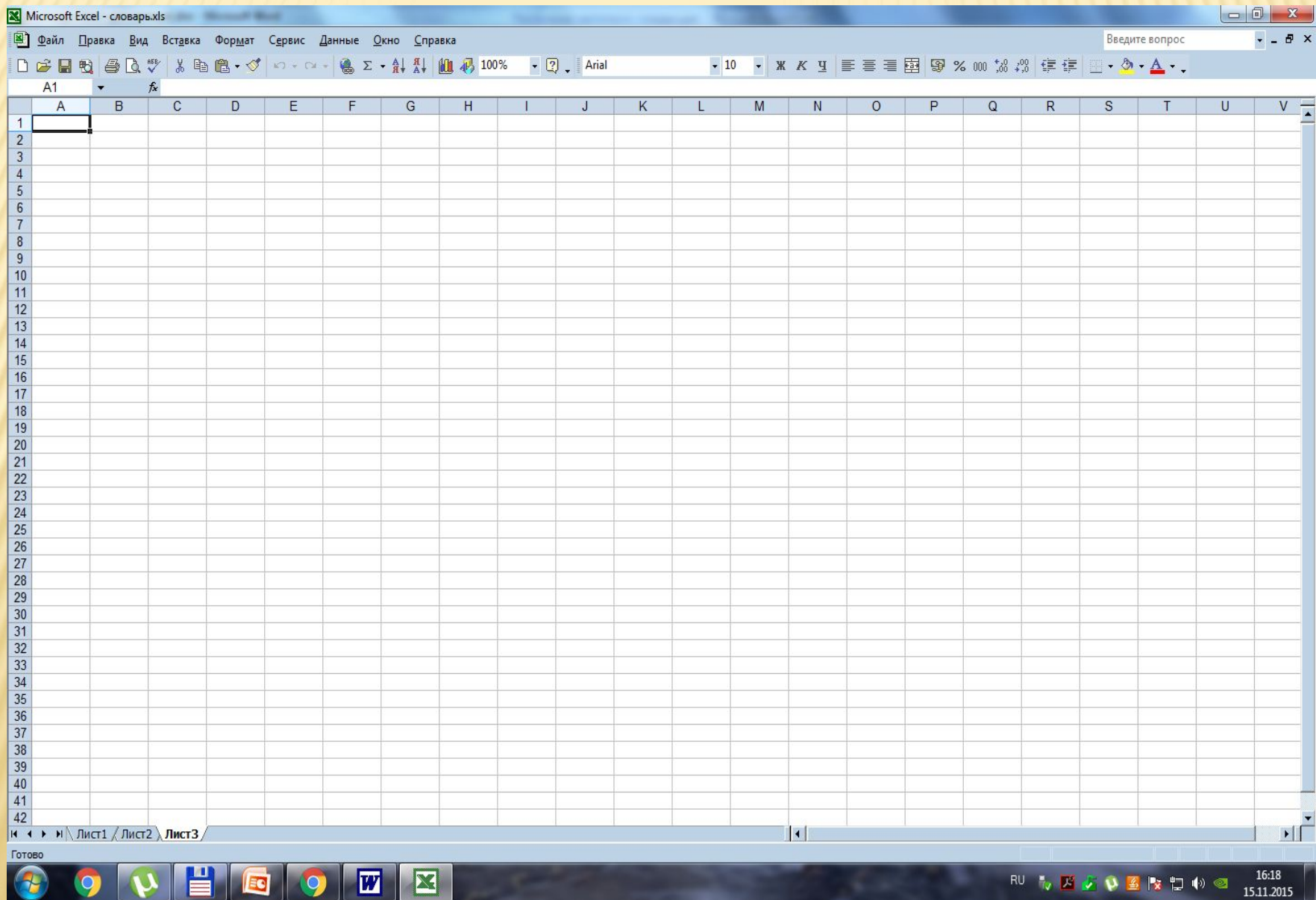
3. Далее ориентируем все слова по центру листа (Рис. 5).



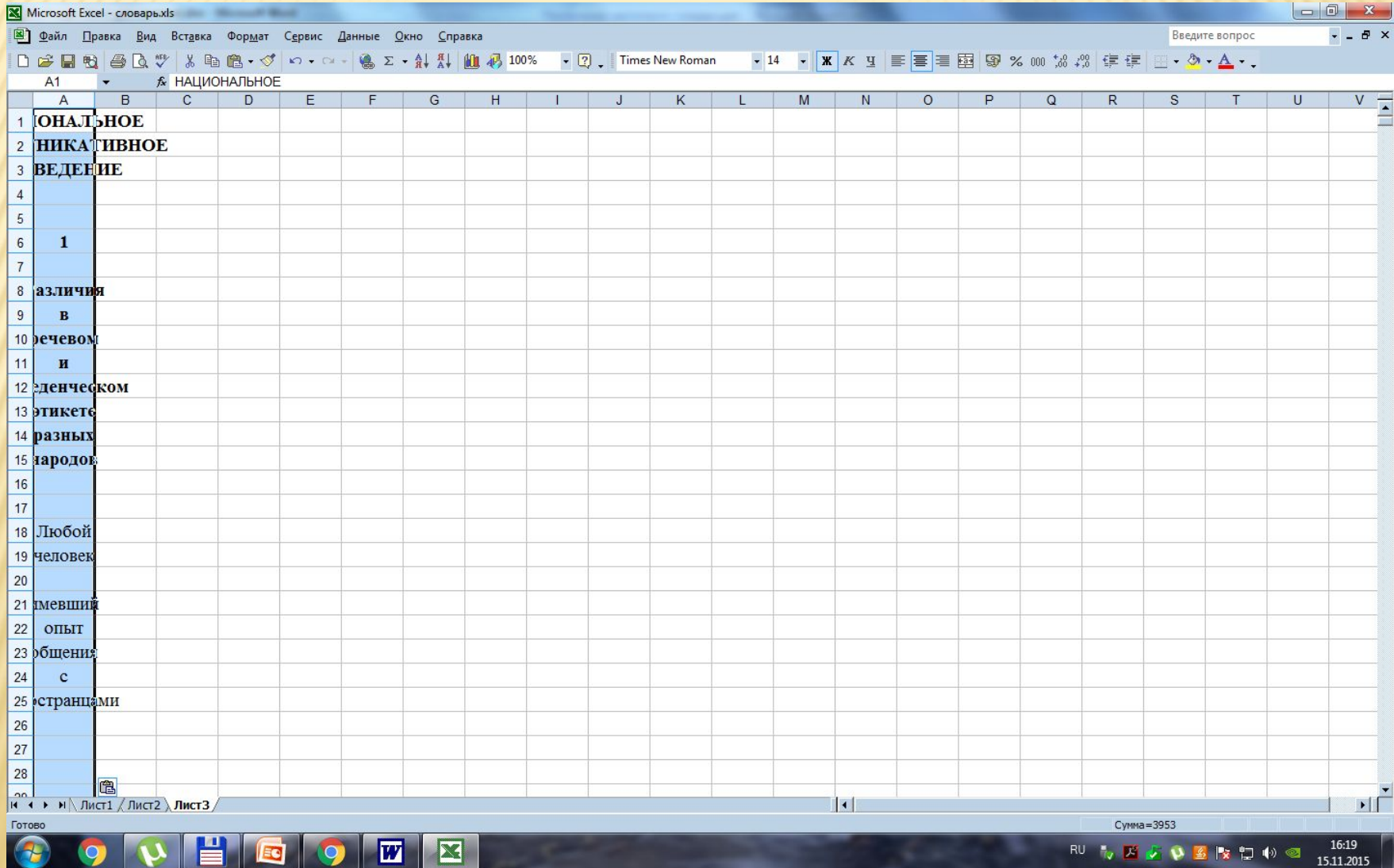
4. С помощью сочетания клавиш Shift и F3 уберем прописные буквы (Рис. 6).



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



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

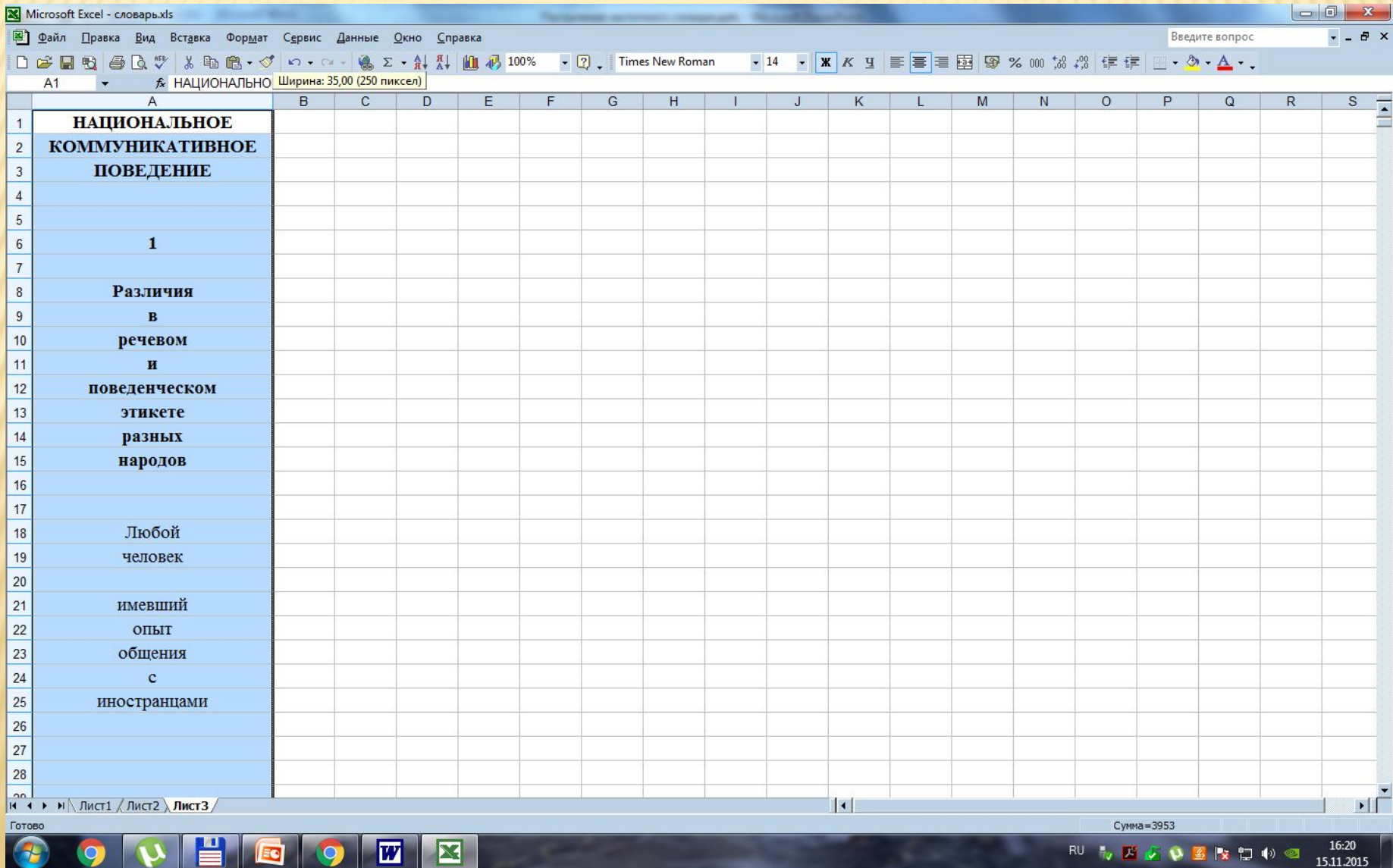


The screenshot shows a Microsoft Excel window titled "Microsoft Excel - словарь.xls". The interface includes a menu bar (Файл, Правка, Вид, Вставка, Формат, Сервис, Данные, Окно, Справка), a ribbon with various tools, and a search bar containing "Введите вопрос". The spreadsheet grid shows column A containing the following text from row 1 to 25:

Row	Column A
1	ОНАЛЬНОЕ
2	НИКАТИВНОЕ
3	ВЕДЕНИЕ
4	
5	
6	1
7	
8	азличия
9	в
10	речевом
11	и
12	денческом
13	этикет
14	разных
15	тародов
16	
17	
18	Любой
19	человек
20	
21	имевший
22	опыт
23	общения
24	с
25	странцами

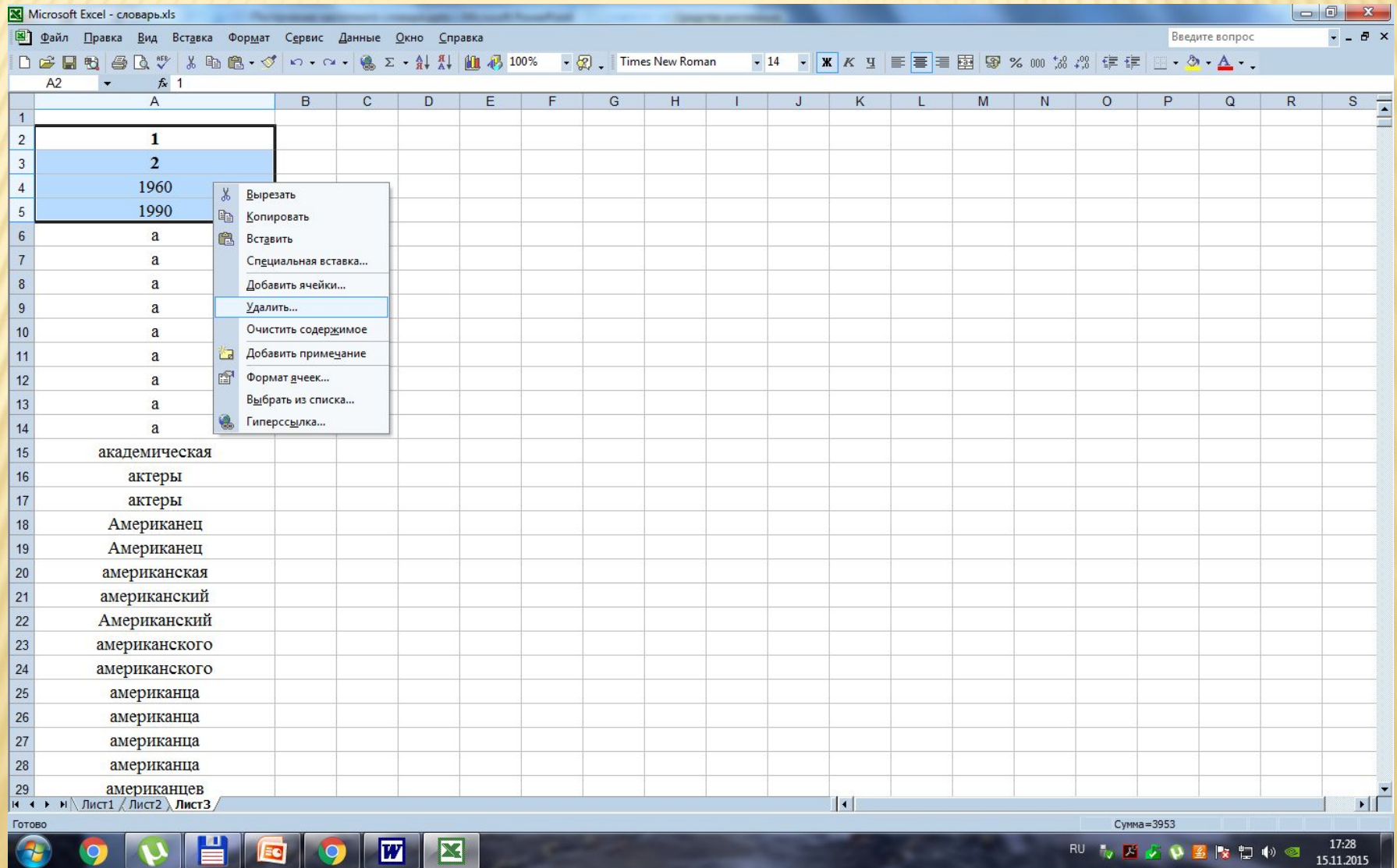
The status bar at the bottom shows "Готово", "Сумма=3953", and the system tray with the date "15.11.2015" and time "16:19".

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

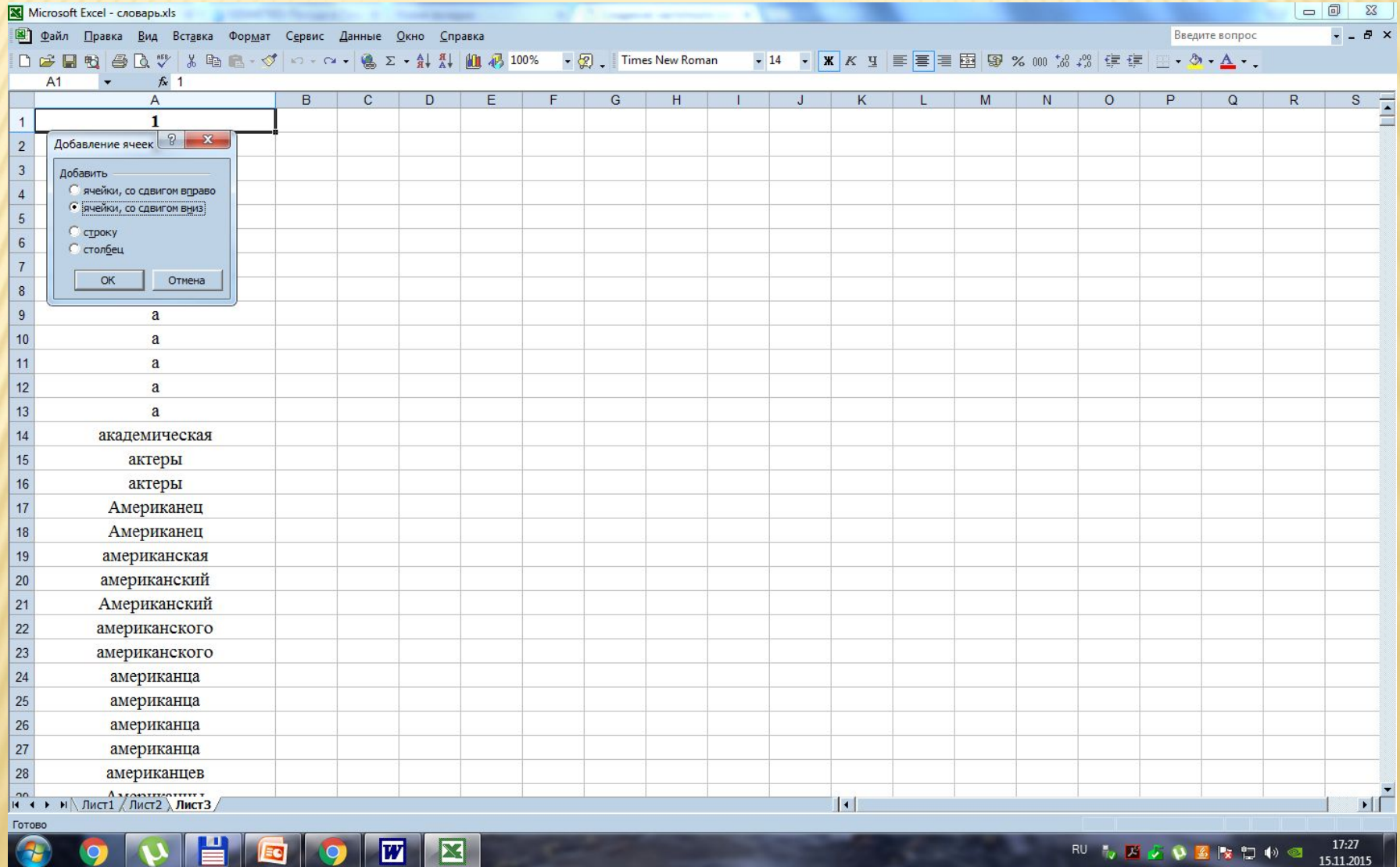


8. Далее сортируем список слов в таблице по алфавиту. Для этого необходимо выделить их нажатием на колонку «А». Затем использовать опцию сортировка по возрастанию (от А до Я). (Рис. 10)

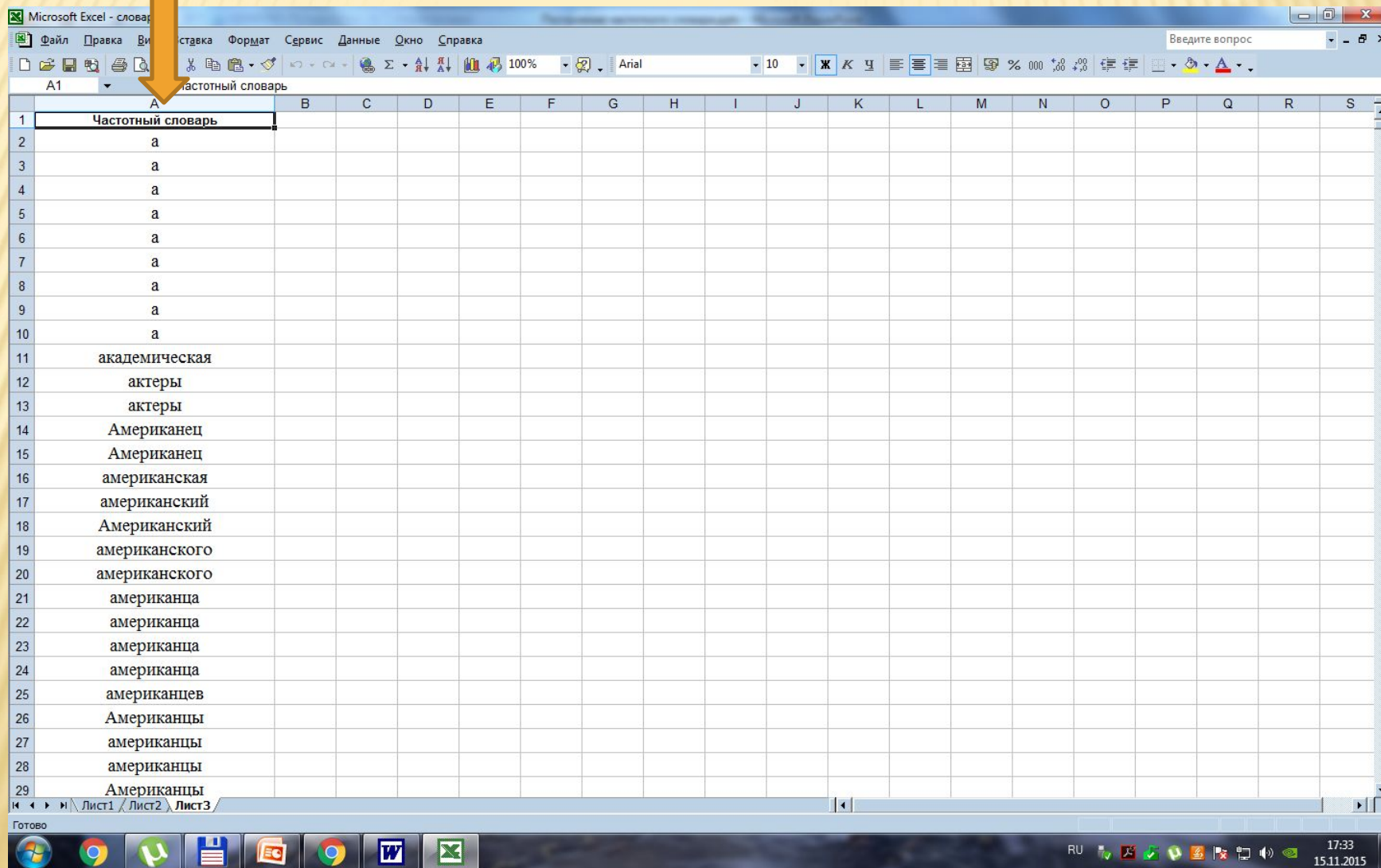
9. При необходимости можно удалить ячейки с цифровыми данными (Рис. 11).



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



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

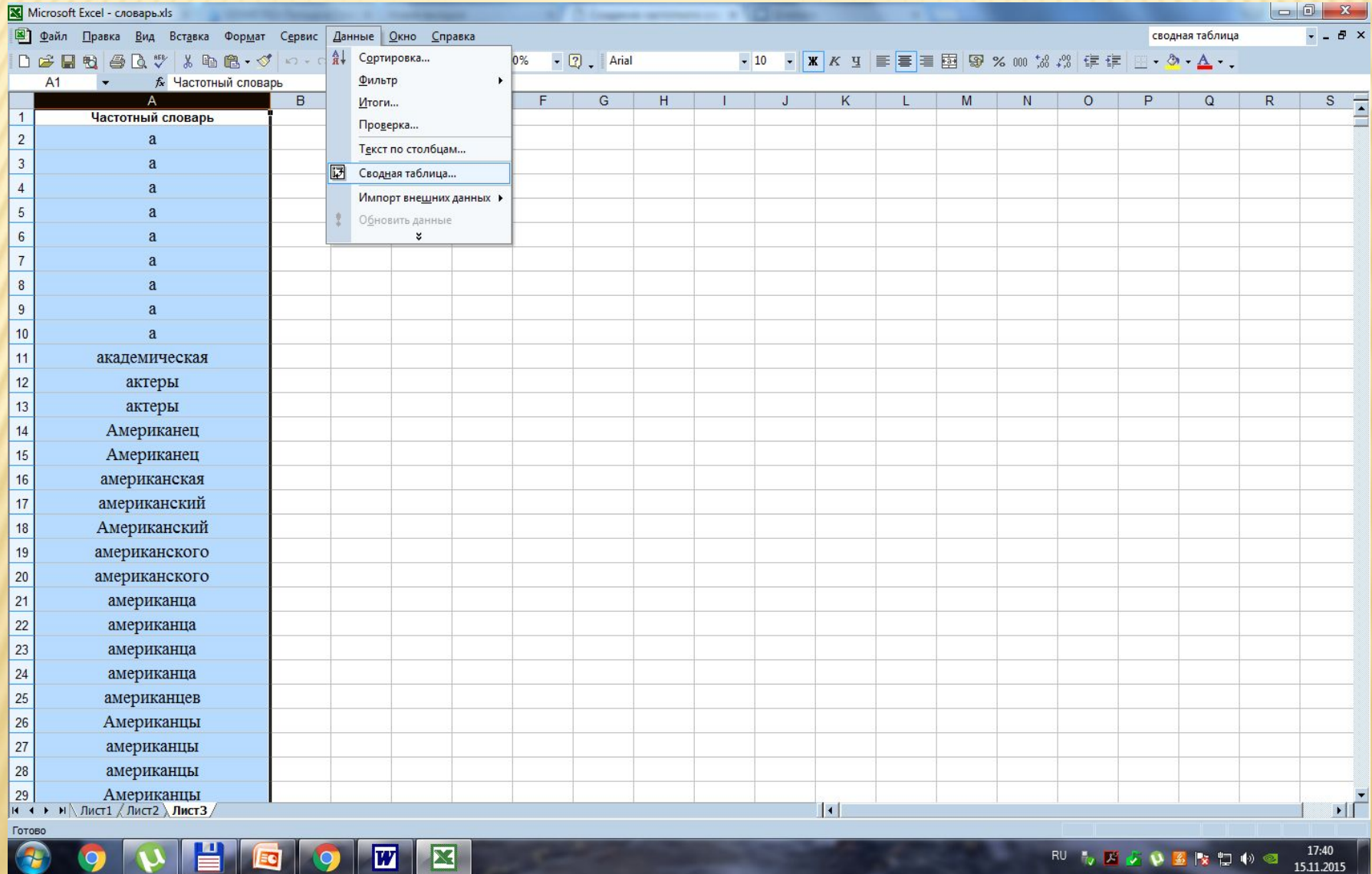


The screenshot shows a Microsoft Excel spreadsheet with a table named "Частотный словарь" starting at cell A1. The table contains the following words:

Row	Word
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

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.



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

Рис. 15

The image shows a Microsoft Excel window titled "Microsoft Excel - словарь.xls". The spreadsheet contains a list of words in column A, with the first cell (A1) containing "Частотный словарь". The words listed are: а, академическая, актеры, актеры, Американец, Американец, американская, американский, Американский, американского, американского, американца, американца, американца, американца, американцев, Американцы, американцы, американцы, and Американцы.

A "Master of PivotTables and Charts" dialog box is open, titled "Мастер сводных таблиц и диаграмм - шаг 1 из 3". It contains the following options:

Создать таблицу на основе данных, находящихся:

- в списке или базе данных Microsoft Excel
- во внешнем источнике данных
- в нескольких диапазонах консолидации
- в другой сводной таблице или сводной диаграмме

Вид создаваемого отчета:

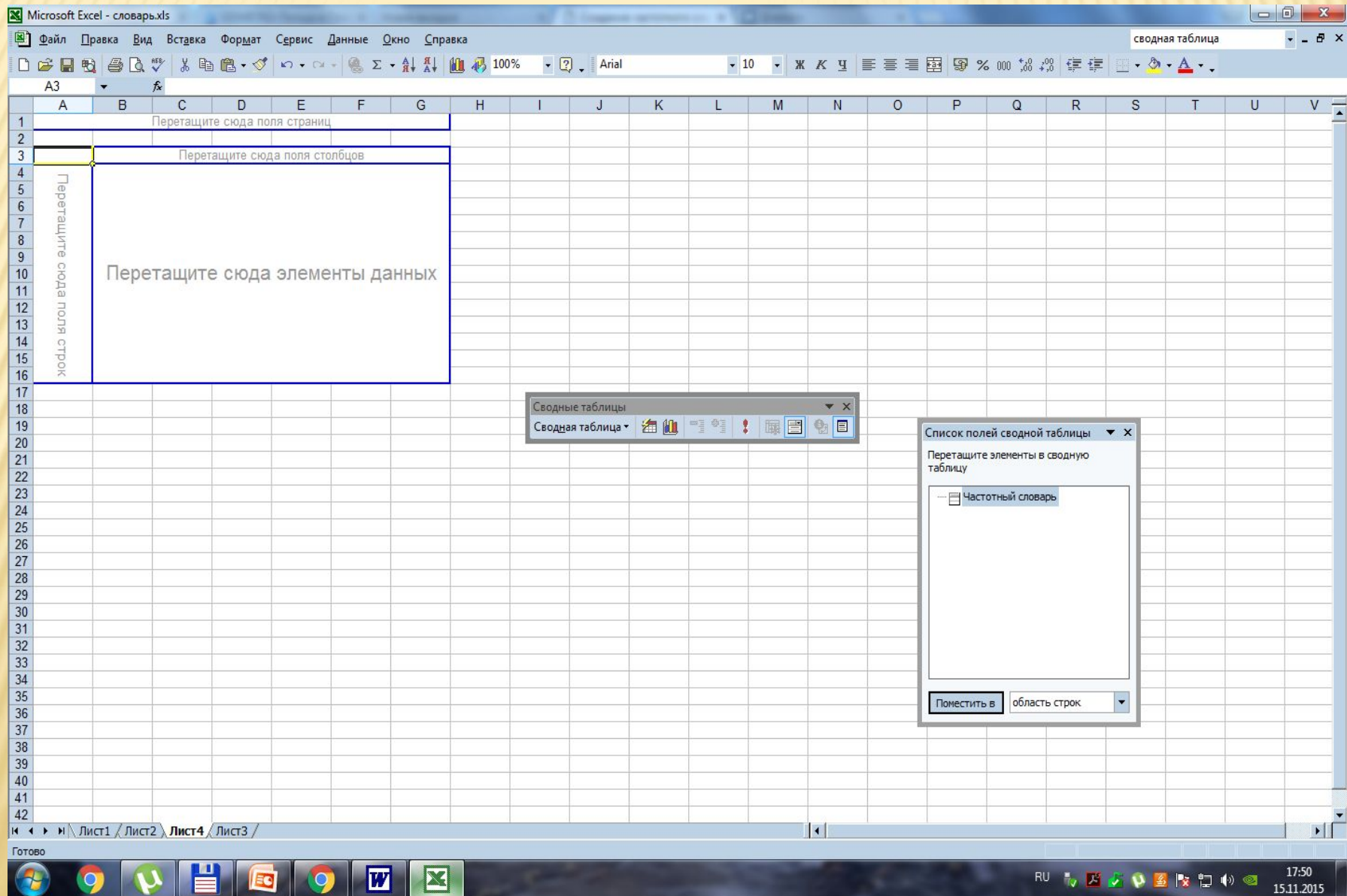
- сводная таблица
- сводная диаграмма (со сводной таблицей)

Buttons at the bottom: Отмена, < Назад, Далее >, Готово.

Рис. 16.

The image shows a screenshot of Microsoft Excel with a PivotTable wizard dialog box open. The spreadsheet contains a list of words in column A, starting with 'Частотный словарь' in row 1 and ending with 'Американцы' in row 29. The PivotTable wizard is at step 3 of 3, titled 'Мастер сводных таблиц и диаграмм - шаг 3 из 3'. It asks 'Понести таблицу в:' (Move table to:) and has two radio button options: 'новый лист' (new sheet) and 'существующий лист' (existing sheet). The 'новый лист' option is selected. Below the options is a text box for specifying the sheet name. At the bottom of the dialog, there are buttons for 'Макет...' (Layout...), 'Параметры...' (Options...), 'Отмена' (Cancel), '< Назад' (Back), 'Далее >' (Next), and 'Готово' (Finish). The status bar at the bottom of the Excel window shows 'Готово' (Finish) and the sheet names 'Лист1', 'Лист2', and 'Лист3'. The system tray at the bottom right shows the date and time: '17:47 15.11.2015'.

Рис. 17.



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.

The image shows a screenshot of Microsoft Excel with a PivotTable. The PivotTable is located in the range B3:G42. The source data is a list of words in column A, starting with 'а' and ending with 'было'. The PivotTable is currently empty, with the following text prompts:

- Row 1: Перетащите сюда поля страниц
- Row 2: (empty)
- Row 3: Перетащите сюда поля столбцов
- Row 4: Перетащите сюда элементы данных

A dialog box titled 'Список полей сводной таблицы' (PivotTable Field List) is open, showing the 'Частотный словарь' (Frequency Dictionary) field. The dialog has a 'Поместить в' (Place in) button and a dropdown menu set to 'область строк' (Row Labels).

An orange arrow points from the top of the dialog box down to the 'Частотный словарь' field in the dialog.

The Excel interface includes the ribbon (Файл, Правка, Вид, Вставка, Формат, Сервис, Данные, Окно, Справка), the formula bar, and the taskbar at the bottom showing the Windows Start button and various application icons.

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

100% Arial 10 Ж К Ч

1 Перетащите сюда поля страниц

2

3

4 Частотный словарь Перетащите сюда поля столбцов

5 а

6 академическая

7 актеры

8 Американец

9 американская

10 американский

11 американского

12 американца

13 американцев

14 Американцы

15 Америкой

16 антрополог

17 аплодировать

18 аплодируют

19 аплодисментами

20 аплодисменты

21 аудитории

22 бегают

23 без

24 белых

25 бельгийцев

26 бесцеремонно

27 благодарить

28 благодарят

29 близких

30 близко

31 Больше

32 Большие

33 большинства

34 большинство

35 брап

36 броситься

37 будто

38 букет

39 бы

40 бывало

41 бывают

42 было

Перетащите сюда элементы данных

Сводные таблицы

Сводная таблица

Список полей сводной таблицы

Перетащите элементы в сводную табл

Частотный словарь

Поместить в область строк

Лист1 Лист2 Лист4 Лист3

Чтобы сделать поле полей данных, отпустите кнопку мыши

RU 18:03 15.11.2015

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

The screenshot shows a Microsoft Excel window titled "словарь.xls". The pivot table is structured as follows:

Частотный словарь	Итого
а	9
академическая	1
актеры	2
Американец	2
американская	1
американский	2
американского	2
американца	4
американцев	1
Американцы	4
Америкой	1
антрополог	1
аплодировать	1
аплодируют	2
аплодисментами	2
аплодисменты	1
аудитории	1
бегают	1
без	1
белых	1
бельгийцев	1
бесцеремонно	1
благодарить	2
благодарят	2
близких	1
близко	3
Больше	1
Большие	2
большинства	1
большинстве	1
брал	1
броситься	1
будто	1
букет	1
бы	1
бывало	1
бывают	1
было	2

Two dialog boxes are open:

- Сводные таблицы**: Shows "Сводная таблица" selected.
- Список полей сводной таблицы**: Shows "Частотный словарь" in the list of fields. The "Поместить в" (Place in) dropdown is set to "область строк" (Rows).

An orange arrow points to the "Итого" row in column B, indicating the frequency column.

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

The screenshot shows a Microsoft Excel window titled "словарь.xls". The spreadsheet contains a list of words in column A and their frequency counts in column B. A "Сортировка" (Sort) dialog box is open, showing the selected range as "\$B\$762" and the sort order set to "по убыванию" (descending). The dialog also includes options for sorting by values or labels and a summary section.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
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	этикетке	1																	
755	этим	1																	
756	этому	1																	
757	является	1																	
758	Японии	1																	
759	японские	1																	
760	японского	1																	
761	японца	1																	
762	(пусто)	1																	
763	Общий итог	1385																	
764																			
765																			
766																			
767																			
768																			
769																			
770																			
771																			
772																			
773																			
774																			
775																			
776																			
777																			
778																			
779																			
780																			
781																			
782																			
783																			

Сортировка

Чтобы сортировка выполнялась автоматически при каждом обновлении отчета, воспользуйтесь командой "Сортировка и лучшая десятка" меню сводной таблицы.

Сортировать по
\$B\$762

по возрастанию
 по убыванию

Сортировать
 значения
 подписи

Сводка
Сортировка Частотный словарь по убыванию по Количеству по полю Частотный словарь

Параметры... OK Отмена

Готово Сумма=1385

18:23 15.11.2015

