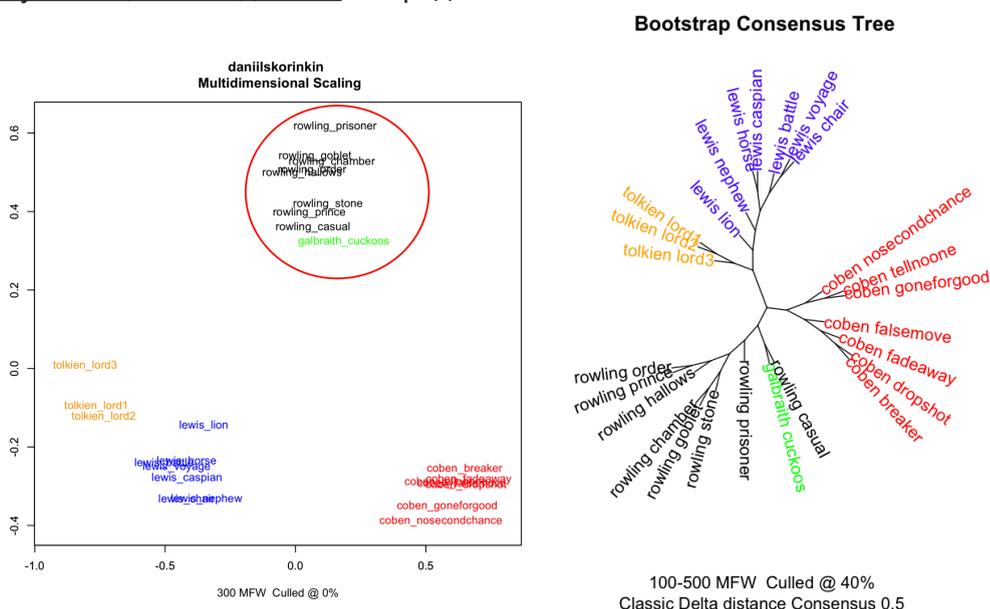


Итак, вам удалось добиться работы stylo. Теперь давайте учиться им пользоваться

Основная функция программы, которой все пользуются, так и называется — *stylo*. Она запускается так (можно пока не запускать):

```
stylo()
```

Эта функция умеет делить тексты в корпусе на слова (вернее, на токены, но сейчас для нас принципиальной разницы нет), считать частотности, вычислять близость текстов на основе Delta и нескольких других метрик — и выдавать результат как в виде таблиц с частотностями и показателями «стилистической близости», так и в виде разных визуализаций этих данных — вроде таких:

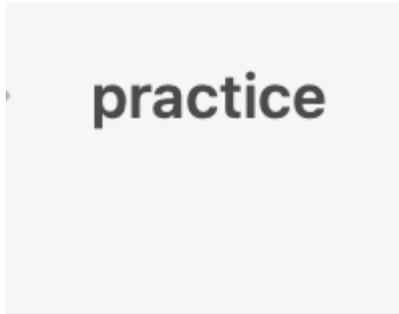


От нас требуется только сообщить программе, откуда брать тексты!

1. Создаем папку для экспериментов

Самая важная вещь, которую нужно знать: **рабочая папка для экспериментов Stylo должна содержать внутри папку под названием corpus. Без этого ничего работать не будет!** Раведите папку под любым названием (желательно без пробелов) а потом **руками в ней создайте папку corpus** (еще раз: это важно, название подпапки должно быть именно **corpus**

То есть где-то на вашем компьютере должна быть папка (например у меня она называется practice) с папкой **corpus** внутри:



corpus

В неё мы уже будем складывать тексты, то есть рабочая папка должна содержать в себе папку **corpus** в которой должны лежать какие-то тексты в самом примитивном формате .txt (прямо в ней, без под-папок, без zip архивов):

< > practice

Name

✓ corpus

 Dostoevsky_BednieLyudi.txt

 Dostoevsky_Besy.txt

 Dostoevsky_PrestuplenieINakazanie.txt

 Dostoevsky_UnijennieIOskorblennie.txt

 Goncharov_Oblomov.txt

 Goncharov_Obryv.txt

 Goncharov_ObyknovennayaIstoriya.txt

 Tolstoy_AnnaKarenina.txt

 Tolstoy_SemeynoeShastye.txt

 Tolstoy_VoinaIMir.txt

 Tolstoy_Voskresenie.txt

 Turgenev_DvoryanskoeGnezdo.txt

 Turgenev_Dym.txt

 Turgenev_OtsylIDeti.txt

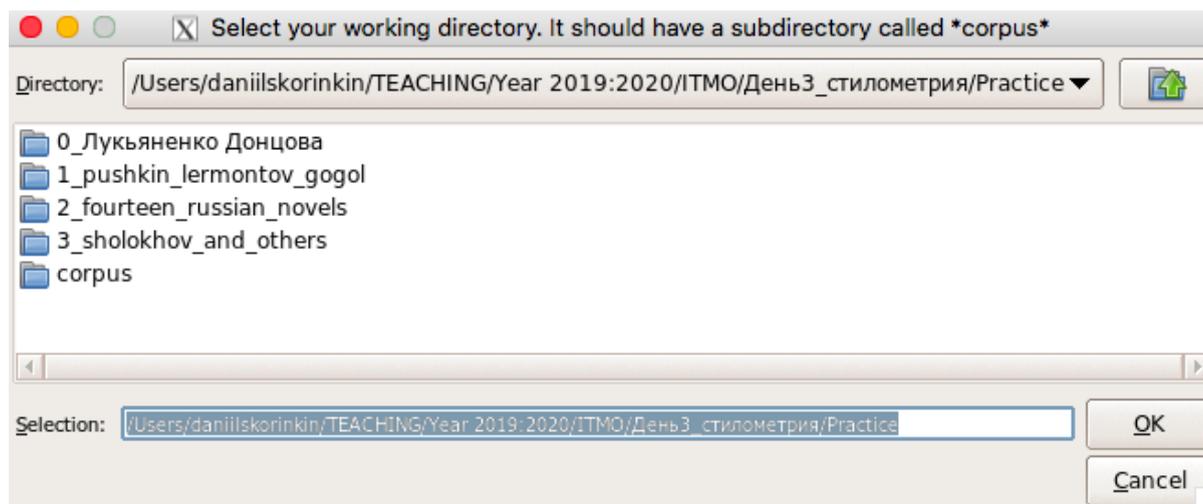
2. Как теперь указать путь к рабочей папке с папкой корпус

Новый способ (работает с 2019 года):

В последних версиях stylo при запуске функции stylo() и нажатии кнопки ок тут



графический интерфейс сам предлагает вам выбрать папку, автоматически открывая проводник.



Если так случилось — выбирайте нужную вам папку (т.е. вашу папку с папкой corpus).

[Шаг назад: Если у вас **еще нет папки corpus**, то сначала заведите папку под любым названием (желательно без пробелов) или еще как-нибудь, как вам удобно, а потом **руками в ней создаете папку corpus** (еще раз: это важно, название подпапки должно быть именно **corpus**). В этой папке должны лежать тексты.]

Но если интерфейс ничего не предложил, то действуем старым способом:

Старый способ указать на папку с текстами

У вас есть рабочая директория (папка), в которой вы сейчас. Узнать её можно, вбив в R консоли инструкцию:

```
getwd()
```

Например, вы вбили `getwd()` и получили такой адрес: «C:/Users/student/Documents». Или `'/Users/Vasya'` или еще какой-то.

Чтобы `stylo` взял ваши тексты, они должны лежать в текущей рабочей директории в подпапке `'corpus'`. Соответственно, у вас есть два варианта:

1. (плохой простой вариант) Пойти на компьютере в эту самую рабочую папку и создать там себе папку `corpus`. Просто руками создать папку по этому адресу и назвать `corpus`. (NB: в Windows папки в проводнике могут называться по-русски, а реальные пути — английские: чтобы попасть в `Users/student/Documents` надо идти в Пользователи -> `student` -> Мои документы).

Чем плох этот вариант? Тем, что выдача `stylo` (таблицы частотностей, таблицы близости, картинки) будет сохраняться в рабочую папку.

2. (вариант получше) Поменять рабочую директорию с помощью команды `setwd()`. Завести папку, назвать ее, например, `stylometry`. Для простоты можно в текущей рабочей директории.

Теперь пропишем путь к этой папке. Поскольку мы создали ее в текущей рабочей директории, можно прописать путь не целиком от корня, а относительный:

```
setwd('stylometry')
```

Как мы уже помним, `stylo` ищет в этой папке подпапку под названием `corpus`.

Давайте создадим и ее внутри папки `stylometry`. То есть вам нужно пойти в созданную вами папку и руками создать там подпапку `corpus`

А теперь в папку `corpus` можно перенести тексты. Откуда их взять?

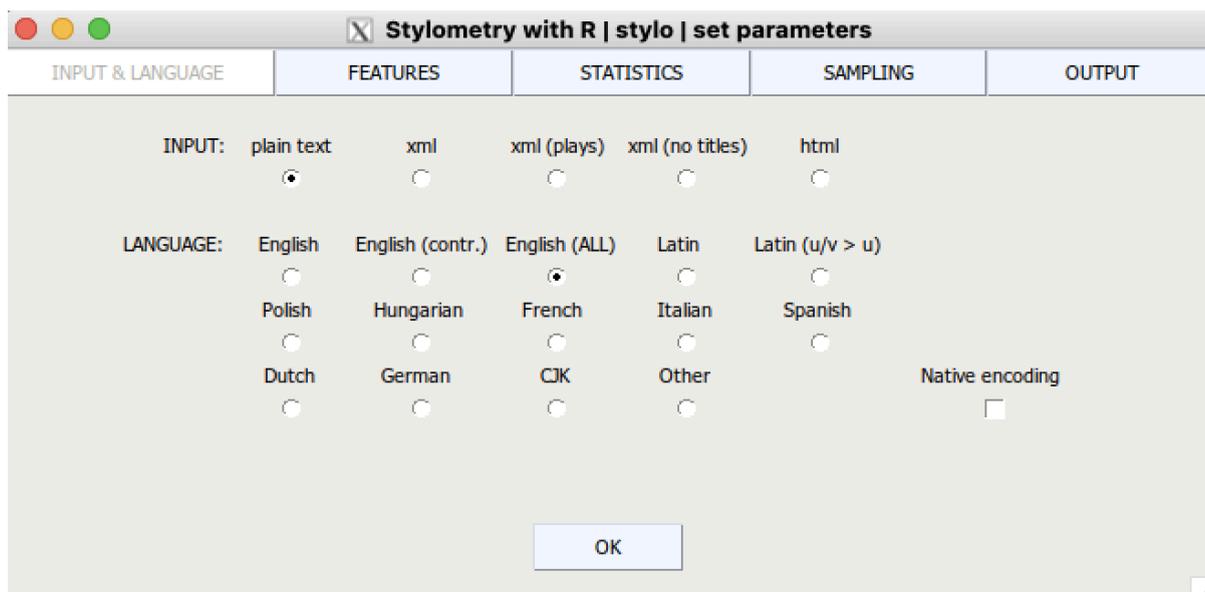
Например, из наших [тренировочных данных](#). Например, возьмем папку `1_pushkin_lermontov_gogol`. Просто копируем все эти текстовые файлы в нашу рабочую папку `corpus`.

Теперь можем запускать `stylo()`:

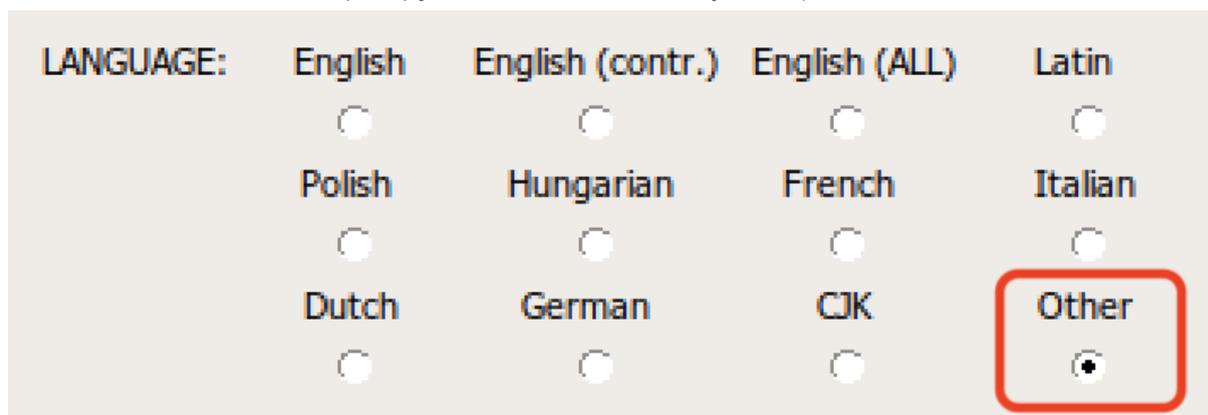
```
stylo()
```

Жмем `enter`

Выскочит такое меню:



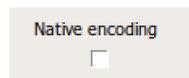
1. В открывшемся меню в первой же вкладке (Input & language) ставим язык «Other». (иначе, если **оставить английский, в текстах учтутся только слова на латинице** и стилометрия русского текста не получится)



2. Если у вас там есть галочка UTF-8, ставим и ее :



(потому что мои файлы — в кодировке Unicode; если не поставить галочку, потом в списках частотностей слов будут кракозябры).

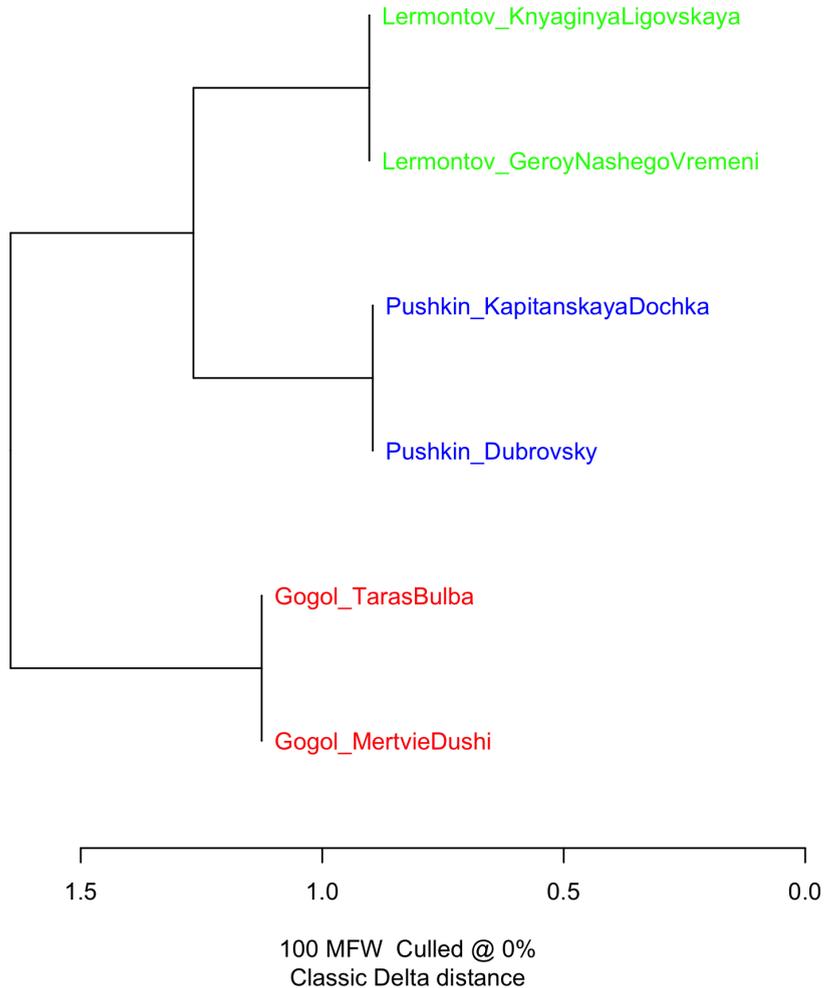


Если вместо UTF-8 галочка «Native encoding», — ничего не делаем.

Остальные настройки пока не трогаем.

Если получилось такое:

stylometry Cluster Analysis



Успех! Вы только что кластеризовали шесть прозаических текстов по их «стилометрическому» расстоянию между ними. Как видите, даже на настройках по умолчанию (когда берутся 100 самых частотных слов в корпусе текстов) Пушкин оказался близок к Пушкину, Лермонтов — Лермонтову, Гоголь — Гоголю.

А вы — научились самой базовой работе со `stylo`. Дальше мы уже будем вместе разбираться на паре.

Ну, и вы можете сами изучать [документацию](#), статьи и [howto](#) и работать самостоятельно.

Некоторые известные баги

Если вы получили:

```
Error in structure(.External(.C_dotTclObjv, objv), class = "tclObj") :  
  [tcl] invalid command name "tk_chooseDirectory".
```

помогает перезапустить RStudio

Если вы Маковод и вам вместо картинки показали ошибку с упоминанием XQuartz -- надо пойти [сюда](#), поставить XQuartz, перезагрузиться и снова по тому же алгоритму.

Если у вас мак/линукс и проблемы с кодировкой — **все слова в тексте превратились в NaN** (посмотреть можно в файлах `wordlist.txt`, `frequency_tables.txt`) — откройте терминал (Terminal) и выполните там эту команду:

```
defaults write org.R-project.R force.LANG en_US.UTF-8
```

Если получили

```
using current directory...  
Error in structure(.External(.C_dotTcl, ...), class = "tclObj") :  
  [tcl] invalid command name "font".
```

попробуйте исполнить:

```
.Tcl("font delete myDefaultFont")
```

и перезапустите RStudio

Бонус: Как запускать функцию `oppose`

Функция `oppose` нужна, чтобы сравнивать два набора текстов и смотреть на то, какие слова часто встречаются в одном и при этом редко — в другом.

Кладем в рабочую директорию `primary_set` и `secondary_set` (готовые есть в [male_female.zip](#))

Потом:

```
oppose(corpus.lang="Other", encoding="UTF-8")
```