Всего я взяла семь языков: английский, французский, испанский, итальянский, румынский, немецкий и датский.

Матрицы для текстов с неубранными стоп-словами и текстами, в которых этих стоп-слов нет, получились очень похожими, разве что для текстов со стоп-словами она чуть менее однородная и отображает пересечения между биграммами французского и испанского языков.

Больше всего пересечений в топах -- между французским и английским языком и испанским и итальянским.

Так как я сделала анализ для очищенных и неочищенных от стоп-слов текстов, то заметила интересное наблюдение: различных общих слов-одиночек оказалось больше в очищенных текстов (видимо, потому, что в неочищенных местах в топах заняли служебные слова).

В целом заметно, что романские языки гораздо ближе друг другу и общие слова в них встречаются чаще, чем в различных представителях германских языков. Довольно много общих слов оказалось для английского и французского. Соответственно, отличить романский язык от неромансского по биграммам было бы довольно просто, как и различить между собой германские языки. С романскими это было бы сделать чуть сложнее. Кроме того, в топе word_scores оказались словосочетания явно романского происхождения, и не всегда можно определить, что к какому языку относится.

Кроме того, я предполагаю, что при анализе были допущены некоторые ошибки: в результате удаления знаков препинания пострадали разного рода диакритические знаки: например, articulo превратилось в artr culo. Очевидно, этот недостаток надо будет устранить при работе с языками в следующий раз.

Приведем конкретные примеры из кода.

Для неочищенных текстов:

Биграмы

```
Итальянский и испанский
{('la presente', 2)}
Испанский и французский
{('que la', 3)}
Датский и немецкий:
{('artikel alle', 3)}
```

Монограмы:

```
Итальянский и испанский:
6
{('momento', 2), ('un', 11), ('proclama', 1), ('barbarie', 1), (
'arbitrariamente', 3), ('base', 2)}

Итальянский и французский:
1
{('barbarie', 1)}
```

```
Испанский и французский:
2
{('ni', 5), ('barbarie', 1)}

Английский и французский:
10
{('race', 2), ('relations', 1), ('religion', 5), ('conscience',
3), ('importance', 1), ('justice', 1), ('nations', 8), ('aspirat
ion', 1), ('respect', 4), ('article', 30)}

Английский и румынский:
1
{('status', 2)}

Датский и английский:
1
{('man', 1)}

Датский и немецкий
1
{('artikel', 30)}
```

Для «чистых» текстов:

Биграммы

```
Английский и французский
{('national international', 2)}
```

Монограммы

```
Итальянский и испанский:
7
{('momento', 2), ('proclama', 1), ('fiduciaria', 1), ('base', 2), ('barbar
ie', 1), ('tal', 1), ('arbitrariamente', 3)}

Итальянский и французский:
1
{('barbarie', 1)}

Французский и испанский:
2
{('barbarie', 1), ('tribunal', 1)}

Французский и английский:
16
{('race', 2), ('importance', 1), ('obligations', 1), ('respect', 4), ('jus
tice', 1), ('nations', 8), ('religion', 5), ('servitude', 1), ('relations'
, 1), ('constitution', 1), ('article', 30), ('impartial', 1), ('torture',
1), ('tribunal', 1), ('aspiration', 1), ('conscience', 3)}

Английский и французский:
1
{('tribunal', 1)}
```

```
Английский и итальянский:
1
{('progressive', 1)}

Датский и немецкий:
1
{('artikel', 30)}
```