Q11

a.
English:
```
$ grep -A 1 '<tuv xml:lang="EN">' uncorpora_plain_20090831.tmx | sed -e 's/<[^>]*>//g' | sed -e 's/--//g'|perl -pe 's/\s/\n/g;'|sort|uniq -c|sort -nr | head -20
```

```
267940 the
175497 of
136607 and
99545 to
66802 in
35910 on
32327 for
22534 that
21181 its
20349 with
20126 a
20006 United
19973 as
17251 by
16995 Nations
13933 at
12827 all
12061 international
11986 States
11173 their
```

Arabic:
```
$ grep -A 1 '<tuv xml:lang="AR">' uncorpora_plain_20090831.tmx | sed -e 's/<[^>]*>//g' | sed -e 's/--//g'|perl -pe 's/\s/\n/g;'|sort|uniq -c|sort -nr | head -20
```

```
92683 في
45238 من
38530 على
34817 -
34466 إلى
24225 أن
18986 التي
18947 وإذ
17776 الأمم
16764 المتحدة
15046 عن
11324 الدول
10979 أو
10175 المؤرخ
9476 كانون
9413 مع
```

9163 جميع
8916 بما
8800 العام
8770 العامة


Spanish:
```
$ grep -A 1 '<tuv xml:lang="ES">' uncorpora_plain_20090831.tmx | sed -
e 's/<[^>]*>//g' | sed -e 's/--//g'|perl -pe 's/\s/\n/g;'|sort|uniq -
c|sort -nr | head -20
```

313375 de
177307 la
131113 y
93456 en
86469 los
82328 el
77855 las
77003 a
69299 que
52778 del
37328 para
28287 con
22750 su
21639 por
21407 al
20266 sobre
18446 Naciones
15557 se
14062 Estados
13952 Unidas

Russian:
```
$ grep -A 1 '<tuv xml:lang="RU">' uncorpora_plain_20090831.tmx | sed -
e 's/<[^>]*>//g' | sed -e 's/--//g'|perl -pe 's/\s/\n/g;'|sort|uniq -
c|sort -nr | head -20
```

135693 и
100876 в
37886 по
37178 на
28031 с
19152 Объединенных
18190 Организации
18148 о
15609 для
14885 от
14180 что
13804 к
13361 Наций

11698 также
10911 года,
9245 декабря
9233 года
8520 их
8379 призывает
7816 или


b.

Top words

| English | AR-EN | ES->EN | RU->EN |
|---|---|---|---|
| the | at | from | and |
| of | From | the | at |
| and | On | Y | by |
| to | — | in | on |
| in | to me | the | from |
| on | that | he | United |
| for | Which | the | The Organization |
| that | Taking | to | about |
| its | Nations | what | for |
| with | United Nations | of the | from |
| a | About | for | what |
| United | Countries | | with | to |
| as | or | his | Of Nations |
| by | Of | by | also |
| Nations | Canon | to the | of the year, |
| at | With | on | December |
| all | all | Nations | of the year |
| international | Including | | HE | their |
| States | General | state | encourages |
| their | the public | | United | or |

In these words, below words are the same in all four languages
AR: at, on, that, nations, United, of, with,all
ES: the, in, to, of, for, with, by, on, nations, state, united
RU: and, at, by, on, united, for, to, nations, of, or

——————————

For the bottom 20
For EN:
1 "22.1
   1 "200"
   1 "100
   1 "10
   1 "(v)
   1 "(p)
   1 "(o)

1 "(n)
    1 "(m)
    1 "(l)
    1 "(k)
    1 "(ix)
    1 "(ii)"
    1 "(i)".
    1 "(e.g.,
    1 "(d)";
    1 "(d)"
    1 "(conclusion
    1 "(b)"
    1 "
For AR:
1 "200"
    1 "2
    1 "11"
    1 "100
    1 "10".
    1 "10
1 "(علــــى
1 "(عقد
1 "(ب)"
1 "(ن )
1 "(م )
1 "(ل )
1 "(ك)
1 "(ع)
1 "(س)
    1 "(3)
    1 "()،
    1 "'9'
    1 "'6'
    1 "'5'
Which means AR->EN
1 "200"
    1 "2
    1 "11"
    1 "100
    1 "10".
    1 "
    1 "(on
    1 "contract
    1 "(b)"
    1 "(n)
    1 "(m)
    1 "(l)
    1 "(k)
    1 "(p)
    1S)

```
    1 "(3)
    (1)
    (I)
    1 "(vi)
    15'

For RU
1 "публичным
    1 "торговля
    1 "ребенок"
    1 "пытками
    1 "выемка"
    1 "судно"
    1 "realize".
    1 "progress
    1 "peacekeeping
    1 "peace
    1 "i)"
    1 "a".
    1 "What
    1 "Visa
    1 "UN
    1 "The
    1 "Reports
    1 "Least
    1 "Info
    1 "Africa
Which means RU->EN
1 "in public
    1 "trade
    1 "child"
    1 "by torture
    1 "notch"
    1 "vessel"
    1 "realize".
    1 "progress
    1 "peacekeeping
    1 "peace
    1 "(i)"
    1 "a".
    1 "What
    1 "Visa
    1 "UN
    1 "The
    1 "Reports
    1 "Least
    1 "Info
    1 "Africa

For ES:
```

```
    1 "8
    1 "600".
    1 "6
    1 "4.5
    1 "3%"
    1 "3
    1 "27C.5.
    1 "23.
    1 "22.1
    1 "2000-2001:
    1 "200"
    1 "2%".
    1 "2
    1 "100%
    1 "10
    1 ",incluso
    1 ",incluida
    1 ",40
    1 "(por
    1 "(concertación
Which means es->EN
1 "8
    1 "600".
    1 "6
    1 "4.5
    1 "3%"
    1 "3
    1 "27C.5.
    1 "23.
    1 "22.1
    1 "2000-2001:
    1 "200"
    1 "2%".
    1 "2
    1 "100%
    1 "10
    1 ", even
    1 "included
    1 ", 40
    1 "(for
    1 "(concertation
```

Conclusion: for the bottom 20, different language has quite different results. For ES,AR, most of the results are numbers. For RU, most are words. For EN, most are single letter.