ZOUHRI YASSINE | DSE

1 - Créez une base de 15 textes à partir du web avec leurs classes respectives : (5 textes de sport, 5 textes d'économie, et 5 textes de politique)

In [1]:
```python
# import statments
import numpy
import pandas as pd
import re
import pprint as pp
from stop_words import get_stop_words
import math
```

In [2]:
```python
# Creation de BE
BE=[["La Francophonie est également présente dans le sport.","sport"],["Au début, le f
t = [b[0] for b in BE ]
tps = [b[1] for b in BE ]
print(tps)
```

```
['sport', 'sport', 'sport', 'sport', 'sport', 'économie', 'économie', 'économie', 'éco
nomie', 'économie', 'politique', 'politique', 'politique', 'politique', 'politique']
```

2 - Créez une fonction qui crée une liste L de tous les mots figurant dans cette base (sans répétition).

In [3]:
```python
# Creation de la liste L de tous les mots figurant dans cette base (sans répétition) e
def creationL(liste_s):
    words = []
    for s in liste_s:
        w = extraction(s)
        words.extend(w)
    words = sorted(list(set(words)))
    return words

def extraction(s):
    punctuation= '''!()-–[]{};:'"\, <>./?@#$%^&*_~«»'''
    words = re.sub("^[a-zA-Z]+", " ",  s).split()
    cleaned_text = []
    for w in words:
        w_new = ''
        for x in w:
            if x not in punctuation and not x.isnumeric():
                w_new = w_new + x
        if len(w_new) == 0:
            continue
        cleaned_text.append(w_new.lower())
    return cleaned_text


vocab = creationL(t)
old_len_vocab = len(vocab)
print("{0} words in this List for the '15 textes'\n \n{1} \n".format(len(vocab), vocab
```

```
119 words in this List for the '15 textes'

['actuel', 'aiment', 'attendue', 'autres', 'aux', 'balkans', 'basketball', 'beaux', 'b
onne', 'budget', 'budgétaire', 'canadienne', 'cardio', 'cest', 'chose', 'comme', 'comm
une', 'comprend', 'compétent', 'conférence', 'consacrer', 'conscience', 'conservatric
e', 'considéré', 'considérés', 'continuent', 'dans', 'de', 'des', 'deuros', 'devriezvo
us', 'dici', 'divisions', 'donc', 'dune', 'débat', 'début', 'démocratie', 'effectif',
'en', 'est', 'et', 'européenne', 'fiscale', 'football', 'forment', 'francophonie', 'gr
andes', 'gérées', 'instruit', 'internationale', 'jeune', 'jours', 'la', 'laisser', 'l
e', 'league', 'leur', 'lexcellent', 'libéraliser', 'lobjectif', 'lon', 'léconomie', 'm
aldives', 'mardi', 'mercredi', 'millions', 'moment', 'mondiales', 'notamment', 'nouvea
u', 'par', 'parvenu', 'petit', 'pleine', 'politiciens', 'politique', 'porte', 'portori
caine', 'pour', 'pourraient', 'preuve', 'professionnelles', 'prospérité', 'présente',
'puissances', 'que', 'repose', 'réduction', 'résultat', 'second', 'secteur', 'service
s', 'simulateur', 'soit', 'soldat', 'sont', 'sport', 'sports', 'stimulerait', 'succè
s', 'suite', 'sur', 'tant', 'terminés', 'the', 'tourisme', 'transition', 'trois', 'u
n', 'une', 'vote', 'à', 'économie', 'économique', 'également', 'émissions', 'épasser',
'être']
```

3 - Créez une première fonction T2V qui transforme chaque texte T en un vecteur d'occurrences des mots de la liste L.

In [4]:
```python
def vector( v, s):
    words = extraction(s)
    bag_vector = numpy.zeros(len(v))
    for w in words:
        for i,word in enumerate(v):
            if word == w:
                bag_vector[i] += 1
    return list(bag_vector)

def t2v(alls, v):
    bags = []
    for s in alls:
        bag_vector = vector(v, s)
        print("\n{0}\n{1}".format(s,numpy.array(bag_vector)))
        bags.append(bag_vector)
    return bags
bg = t2v(t, vocab)
```

```
La Francophonie est également présente dans le sport.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.]

Au début, le football est considéré comme un petit moment économique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0.]

Les sports aiment le basket-ball, le football, le football pourraient être considérés
cardio.
[0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 2. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 3. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]

FIFA 14 est la suite tant attendue de l'excellent simulateur de football.
[0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 2. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les trois autres divisions professionnelles sont gérées par The football League.
[0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les beaux jours de l'économie conservatrice sont terminés, pour le moment.
[0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les grandes puissances mondiales continuent de libéraliser leur économie.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]

Les services forment le second secteur de l'économie portoricaine et comprend notammen
t le tourisme.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 2. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

La prospérité de l'économie canadienne repose sur un effectif compétent et instruit.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Dépasser l'objectif actuel de réduction de 20 % des émissions d'ici 2020 stimulerait
l'économie européenne (débat mardi, vote mercredi).
[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 2. 1. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 1. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 1. 0.]

Peut être devriez-vous laisser la politique aux politiciens, soldat.
[0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]

Les Maldives sont une jeune démocratie, en pleine transition politique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]

Votre conférence porte sur la politique budgétaire et fiscale internationale.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Fais donc preuve de conscience politique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Ce succès est le résultat d'une politique commune, et c'est une bonne chose que l'on s
oit à nouveau parvenu à consacrer 839 millions d'euros aux Balkans dans le budget 200
1.
[0. 0. 0. 0. 1. 1. 0. 0. 1. 1. 0. 0. 0. 1. 1. 0. 1. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 2. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0.
 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 1. 0.
 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 2. 0. 0. 0. 0. 0. 0.]
```

In [5]:
```python
df_t2v = pd.DataFrame(bg, columns=vocab)
df_t2v
```

Out[5]:

	actuel	aiment	attendue	autres	aux	balkans	basketball	beaux	bonne	budget	...	un	une	vote	
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	0.0	0.
2	0.0	1.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
3	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
4	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	0.0	0.
9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.
10	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	1.0	0.0	0.
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.
14	0.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	...	0.0	1.0	0.0	2.

15 rows × 119 columns

4 - Créez une liste de mots vides MV (pronoms, prépositions, conjonctions) et utilisez la pour réduite la dimension de L. Quel est le taux de réduction obtenu ?

In [6]:
```python
def mv(vc):
    ignore = get_stop_words('french')
    v = vc
    for x in v:
        if x in ignore:
            v = [temp for temp in v if temp != x ]
    return v
mv_vocab = mv(vocab)
mv_len_vocab = len(mv_vocab)
print("Le taux de reduction est de : {0} %\n".format(1-(mv_len_vocab/old_len_vocab)))
print("{0} words in this List for the '15 textes' \n \n {1} \n".format(len(mv_vocab),
```

```
Le taux de reduction est de : 0.2016806722689075 %

95 words in this List for the '15 textes'

 ['actuel', 'aiment', 'attendue', 'autres', 'balkans', 'basketball', 'beaux', 'bonne',
'budget', 'budgétaire', 'canadienne', 'cardio', 'cest', 'chose', 'commune', 'compren
d', 'compétent', 'conférence', 'consacrer', 'conscience', 'conservatrice', 'considér
é', 'considérés', 'continuent', 'deuros', 'devriezvous', 'dici', 'divisions', 'dune',
'débat', 'démocratie', 'effectif', 'européenne', 'fiscale', 'football', 'forment', 'fr
ancophonie', 'grandes', 'gérées', 'instruit', 'internationale', 'jeune', 'jours', 'lai
sser', 'league', 'lexcellent', 'libéraliser', 'lobjectif', 'lon', 'léconomie', 'maldiv
es', 'mardi', 'mercredi', 'millions', 'moment', 'mondiales', 'notamment', 'parvenu',
'petit', 'pleine', 'politiciens', 'politique', 'porte', 'portoricaine', 'pourraient',
'preuve', 'professionnelles', 'prospérité', 'présente', 'puissances', 'repose', 'réduc
tion', 'résultat', 'second', 'secteur', 'services', 'simulateur', 'soldat', 'sport',
'sports', 'stimulerait', 'succès', 'suite', 'tant', 'terminés', 'the', 'tourisme', 'tr
ansition', 'trois', 'vote', 'économie', 'économique', 'également', 'émissions', 'épass
er']
```

In [7]:
```python
bg_mv = t2v(t, mv_vocab)
```

```
La Francophonie est également présente dans le sport.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.]

Au début, le football est considéré comme un petit moment économique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.]

Les sports aiment le basket-ball, le football, le football pourraient être considérés
cardio.
[0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 2. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

FIFA 14 est la suite tant attendue de l'excellent simulateur de football.
[0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les trois autres divisions professionnelles sont gérées par The football League.
[0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0.]

Les beaux jours de l'économie conservatrice sont terminés, pour le moment.
[0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les grandes puissances mondiales continuent de libéraliser leur économie.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]

Les services forment le second secteur de l'économie portoricaine et comprend notammen
t le tourisme.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]

La prospérité de l'économie canadienne repose sur un effectif compétent et instruit.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Dépasser l'objectif actuel de réduction de 20 % des émissions d'ici 2020 stimulerait
l'économie européenne (débat mardi, vote mercredi).
[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 1. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 1.]

Peut être devriez-vous laisser la politique aux politiciens, soldat.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les Maldives sont une jeune démocratie, en pleine transition politique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0.]

Votre conférence porte sur la politique budgétaire et fiscale internationale.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Fais donc preuve de conscience politique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Ce succès est le résultat d'une politique commune, et c'est une bonne chose que l'on s
oit à nouveau parvenu à consacrer 839 millions d'euros aux Balkans dans le budget 200
1.
[0. 0. 0. 0. 1. 0. 0. 1. 1. 0. 0. 0. 1. 1. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
```

In [8]:
```python
df_mv = pd.DataFrame(bg_mv, columns=mv_vocab)
df_mv
```

Out[8]:

	actuel	aiment	attendue	autres	balkans	basketball	beaux	bonne	budget	budgétaire	...	the
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
2	0.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	...	0.0
3	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
4	0.0	0.0	0.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	1.0
5	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0	...	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
9	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0
14	0.0	0.0	0.0	0.0	1.0	0.0	0.0	1.0	1.0	0.0	...	0.0

15 rows × 95 columns

5 - Créez une fonction ELIM qui élimine les mots de la liste L ayant figuré dans les 3 classes et garde uniquement ceux qui figurent dans une ou deux classes. Quel est le taux de réduction obtenu ?

In [9]:
```python
def checker(temp):
    for t in temp:
        if t > numpy.float64(0):

            return True
    return False
def elim(dframe):
    z =  numpy.zeros(3)
    mots = []
    for c in dframe.columns.to_list():
        s = dframe[c][0:5].to_list()
        e = dframe[c][5:10].to_list()
        p = dframe[c][10:].to_list()
        if checker(s) and checker(e) and checker(p):
            mots.append(c)
    print("Les mots de la liste L, ayant figuré dans les 3 classes, sont :\n {0}".form
    return mots
print("Apres T2V : \n")
elim = elim(df_t2v)
```

```
Apres T2V :

Les mots de la liste L, ayant figuré dans les 3 classes, sont :
 ['de', 'le', 'sont']
```

In [10]:
```python
elim_vocab = [temp for temp in vocab if temp not in elim ]
elim_len_vocab = len(elim_vocab)

print("Le taux de reduction est de : {0}\n".format(1-(elim_len_vocab/old_len_vocab)))
print("{0} words in this List for the '15 textes' \n \n {1} \n".format(len(elim_vocab
bg_mv_elim = t2v(t, elim_vocab)
```

```
Le taux de reduction est de : 0.025210084033613467

116 words in this List for the '15 textes'

 ['actuel', 'aiment', 'attendue', 'autres', 'aux', 'balkans', 'basketball', 'beaux',
'bonne', 'budget', 'budgétaire', 'canadienne', 'cardio', 'cest', 'chose', 'comme', 'co
mmune', 'comprend', 'compétent', 'conférence', 'consacrer', 'conscience', 'conservatri
ce', 'considéré', 'considérés', 'continuent', 'dans', 'des', 'deuros', 'devriezvous',
'dici', 'divisions', 'donc', 'dune', 'débat', 'début', 'démocratie', 'effectif', 'en',
'est', 'et', 'européenne', 'fiscale', 'football', 'forment', 'francophonie', 'grande
s', 'gérées', 'instruit', 'internationale', 'jeune', 'jours', 'la', 'laisser', 'leagu
e', 'leur', 'lexcellent', 'libéraliser', 'lobjectif', 'lon', 'léconomie', 'maldives',
'mardi', 'mercredi', 'millions', 'moment', 'mondiales', 'notamment', 'nouveau', 'par',
'parvenu', 'petit', 'pleine', 'politiciens', 'politique', 'porte', 'portoricaine', 'po
ur', 'pourraient', 'preuve', 'professionnelles', 'prospérité', 'présente', 'puissance
s', 'que', 'repose', 'réduction', 'résultat', 'second', 'secteur', 'services', 'simula
teur', 'soit', 'soldat', 'sport', 'sports', 'stimulerait', 'succès', 'suite', 'sur',
'tant', 'terminés', 'the', 'tourisme', 'transition', 'trois', 'un', 'une', 'vote',
'à', 'économie', 'économique', 'également', 'émissions', 'épasser', 'être']


La Francophonie est également présente dans le sport.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.]

Au début, le football est considéré comme un petit moment économique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0.]

Les sports aiment le basket-ball, le football, le football pourraient être considérés
cardio.
[0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 2. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]

FIFA 14 est la suite tant attendue de l'excellent simulateur de football.
[0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les trois autres divisions professionnelles sont gérées par The football League.
[0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les beaux jours de l'économie conservatrice sont terminés, pour le moment.
[0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Les grandes puissances mondiales continuent de libéraliser leur économie.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.]

Les services forment le second secteur de l'économie portoricaine et comprend notammen
t le tourisme.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 1. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

La prospérité de l'économie canadienne repose sur un effectif compétent et instruit.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Dépasser l'objectif actuel de réduction de 20 % des émissions d'ici 2020 stimulerait
l'économie européenne (débat mardi, vote mercredi).
[1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 1. 0.]

Peut être devriez-vous laisser la politique aux politiciens, soldat.
[0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1.]

Les Maldives sont une jeune démocratie, en pleine transition politique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 1. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0.]

Votre conférence porte sur la politique budgétaire et fiscale internationale.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0.
 0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Fais donc preuve de conscience politique.
[0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 1. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0.]

Ce succès est le résultat d'une politique commune, et c'est une bonne chose que l'on s
oit à nouveau parvenu à consacrer 839 millions d'euros aux Balkans dans le budget 200
1.
[0. 0. 0. 0. 1. 1. 0. 0. 1. 1. 0. 0. 0. 1. 1. 0. 1. 0. 0. 0. 1. 0. 0. 0.
 0. 0. 1. 0. 1. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0.
 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 1. 0.
 0. 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0.
 0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 2. 0. 0. 0. 0.]
```

6 - Créez une fonction DIST qui calcule la distance euclidienne entre deux vecteurs

In [11]:
```python
def dist( a, b):
    result = 0
    if len(a) == len(b):
        for i in range(len(a)):
            result += (a[i]-b[i])**2
        result = math.sqrt(result)
    else:
        print("les deux vecteurs n'ont pas la même longueur !!")
    return result
```

7 - Proposez un texte pour lequel : (en utilisant T2V)

- **Parmi les 15 textes de la base, la DIST minimale correspond à la classe économie.**

In [12]:
```python
#définition de KNN
def knn(k,t,B):
    #créer la table des distances t-B
    B = [[b,t] for b,t in zip(B,tps)]
    tD=[]
    for i in range(len(B)):
        tD.append([dist(t,B[i][0]),B[i][1]])
    #trier la tD
    for i in range(len(tD)):
        for j in range(len(tD)):
            if tD[i][0]>tD[j][0]:
                temp=tD[i]
                tD[i]=tD[j]
                tD[j]=temp
    pp.pprint(tD)
    #calculer les k derniers de la tD
    #créer les tables tC et tO
    tC=[]
    tO=[]
    for i in range(len(tD)):
        if not B[i][1] in tC:
            tC.append(B[i][1])
            tO.append(0)

    #remplir la table tO
    for i in range(1,k+1):
        for j in range(len(tC)):
            if tD[-i][1]==tC[j]:
                tO[j]=tO[j]+1
    pp.pprint(tO)
    #retourner la classe prédite
    if tO.count(max(tO)) == 1:
        indmax=tO.index(max(tO))
        return tC[indmax]
    return tD[-1][1]
```

In [13]:
```python
texte_eco = "Les exportations de l'économie grecque sont traditionnellement faibles."
bag_vector_eco = vector(vocab, texte_eco)

predicted_eco = knn(4, bag_vector_eco, bg)
print("\nParmi les 15 textes de la base, la DIST minimale (en utilisant T2V) correspon
```

```
[[5.830951894845301, 'politique'],
 [4.795831523312719, 'sport'],
 [3.872983346207417, 'économie'],
 [3.7416573867739413, 'économie'],
 [3.605551275463989, 'sport'],
 [3.4641016151377544, 'politique'],
 [3.3166247903554, 'politique'],
 [3.3166247903554, 'sport'],
 [3.3166247903554, 'sport'],
 [3.1622776601683795, 'politique'],
 [3.1622776601683795, 'sport'],
 [3.1622776601683795, 'économie'],
 [3.0, 'économie'],
 [2.6457513110645907, 'économie'],
 [2.449489742783178, 'politique']]
[0, 3, 1]

Parmi les 15 textes de la base, la DIST minimale (en utilisant T2V) correspond à la cl
asse : ÉCONOMIE
```

8 - Proposez un texte pour lequel : (en utilisant T2V + MV + ELIM)

- **Parmi les 15 textes de la base, la DIST minimale correspond à la classe sport.**

In [14]:
```python
texte_sport = "Je suis très heureux de la performance du défenseur et de celle de l'éc
bag_vector_sport = vector(mv_vocab, texte_sport)

predicted_sport = knn(4, bag_vector_sport, bg_mv)
print("\nParmi les 15 textes de la base, la DIST minimale (en utilisant T2V + MV + ELI
```

```
[[4.0, 'politique'],
 [3.7416573867739413, 'économie'],
 [3.1622776601683795, 'économie'],
 [2.8284271247461903, 'économie'],
 [2.6457513110645907, 'sport'],
 [2.6457513110645907, 'sport'],
 [2.6457513110645907, 'économie'],
 [2.6457513110645907, 'politique'],
 [2.6457513110645907, 'politique'],
 [2.6457513110645907, 'économie'],
 [2.449489742783178, 'politique'],
 [2.23606797749979, 'sport'],
 [2.23606797749979, 'sport'],
 [2.0, 'politique'],
 [2.0, 'sport']]
[3, 0, 1]

Parmi les 15 textes de la base, la DIST minimale (en utilisant T2V + MV + ELIM) corres
pond à la classe : SPORT
```

9 - Proposez un texte pour lequel :

- **La classe de la DIST minimale (en utilisant T2V) est différente de la classe de la DIST minimale (en utilisant T2V + MV + ELIM)**

In [15]:
```python
texte_9 = "Au début, le football est comme une jeune démocratie qui repose sur la poli
bag_vector_t2v = vector(vocab, texte_9)

predicted_t2v = knn(4, bag_vector_t2v, bg)
print("\nParmi les 15 textes de la base, la DIST minimale (en utilisant T2V)  correspo
```

```
[[5.744562646538029, 'politique'],
 [5.477225575051661, 'économie'],
 [4.795831523312719, 'économie'],
 [4.69041575982343, 'sport'],
 [4.47213595499958, 'sport'],
 [4.47213595499958, 'économie'],
 [4.47213595499958, 'économie'],
 [4.358898943540674, 'économie'],
 [4.242640687119285, 'sport'],
 [4.0, 'politique'],
 [3.872983346207417, 'politique'],
 [3.872983346207417, 'politique'],
 [3.872983346207417, 'sport'],
 [3.605551275463989, 'politique'],
 [3.4641016151377544, 'sport']]
[2, 0, 2]

Parmi les 15 textes de la base, la DIST minimale (en utilisant T2V)  correspond à la c
lasse : SPORT
```

In [16]:
```python
bag_vector_t2v_mv_elim = vector(elim_vocab, texte_9)

predicted_mv_t2v_elim = knn(4, bag_vector_t2v_mv_elim, bg_mv_elim)
print("\nParmi les 15 textes de la base, la DIST minimale (en utilisant T2V + MV + ELI
```

```
[[5.656854249492381, 'politique'],
 [5.0, 'économie'],
 [4.58257569495584, 'économie'],
 [4.242640687119285, 'économie'],
 [4.242640687119285, 'sport'],
 [4.242640687119285, 'sport'],
 [4.242640687119285, 'économie'],
 [4.123105625617661, 'économie'],
 [3.872983346207417, 'politique'],
 [3.872983346207417, 'sport'],
 [3.7416573867739413, 'politique'],
 [3.605551275463989, 'politique'],
 [3.605551275463989, 'sport'],
 [3.4641016151377544, 'sport'],
 [3.3166247903554, 'politique']]
[2, 0, 2]

Parmi les 15 textes de la base, la DIST minimale (en utilisant T2V + MV + ELIM) corres
pond à la classe : POLITIQUE
```

10 - Quel modèle d'apprentissage artificiel avez-vous utilisé dans les questions 7 à 9 ? Donnez 1 avantage et 1 inconvénient de ce modèle.

Le modèle d'apprentissage artificiel utilisé dans les questions 7 à 9 est : KNN

Un avantage :

- Il s'agit d'un modèle simple et facile à appliquer dans la pratique..

Un inconvénient :

- Le modèle est plus lent lorsque le nombre de ses échantillons d'apprentissage est plus élevé..