Lingüística Computacional
Ejercicio 10

1. (10 points) El corpus Brown está etiquetado y analizado por lingüistas. La forma que tiene cuando se descarga es así:

```
For/in the/at first/od time/nn in/in history/nn ,/, the/at U.S./np has/hvz
 produced/vbn a/at society/nn in/in which/wdt less/ap than/in
 one-tenth/nn of/in the/at people/nns turn/vb out/rp so/ql much/ap
 food/nn that/cs the/at Government's/nn$-tl most/ql embarrassing/vbg
  problem/nn is/bez how/wrb to/to dispose/vb inconspicuously/rb
  of/in 100/cd million/cd tons/nns of/in surplus/nn farm/nn produce/nn ./.
```

   NLTK tiene un método (`raw`) que nos da todo el texto del corpus en su forma original. También tiene otro método (`words`) que permite obtener las palabras. Aparentemente es una lista, pero si examinan el tipo les dice que es:

```
<class 'nltk.corpus.reader.util.ConcatenatedCorpusView'>
```

   En realidad las palabras del corpus se pueden obtener mediante una expresión regular. La tarea consiste en escribir una expresión regular para obtener las palabras del corpus a partir del texto original completo obtenido con `raw`.

   Pueden usar las palabras que se obtienen con `words` para comparar el resultado de su expresión regular. Su expresión regular estará bien cuando se obtengan con ella exactamente las palabras que se obtienen con `words`.