1 Introduction

1.1 Preliminaries: Linux tools for handling texts.

1. If you are not familiar with Linux commands such as: grep, sort, uniq, head, tail, cut, paste, gawk, etc. check their manpages and get acquainted with them. You can get a good introduction following the exercises in:

   K.W. Church, *Unix For Poets*
   `http://www.lsi.upc.edu/ padro/Unixforpoets.pdf`

   NOTE: Modern Linux commands may differ in some option flags or parameters from examples used in this tutorial. If you find problems, check the command manpage to find out the right parameters or options.

1.2 Zipf's Laws

1. Write a program to check Zipfs first law ($f = K/r$) on a real corpus: Count word frecuencies, sort them by rank, and plot the curve.

2. Compute the proportionality constant (K) between rank and frequency for each word. Compute its average and deviation. Discuss the results. Are they consistent with Zipfs Law?

NOTE: Use the text files `corpus/en.txt` and `corpus/es.txt`