

De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra

# L'ordinador ens ensenya la llengua



Estàs ben asseguda  
o ben assegut i  
veus bé la  
pantalla?

Ja has marcat la  
teva estatura a la  
columna de caixes  
de cartró en  
entrar?

# Primera pregunta

Com s'ho fan els  
ordinadors per  
ajudar-nos a corregir  
un text?

Els ordinadors  
recorden  
moltíssimes paraules,  
però no saben res  
sobre les paraules que  
recorden.

Podem fer-li revisar  
textos esbojarrats  
però escrits amb  
mots correctes:

## Quina és la frase correcta?

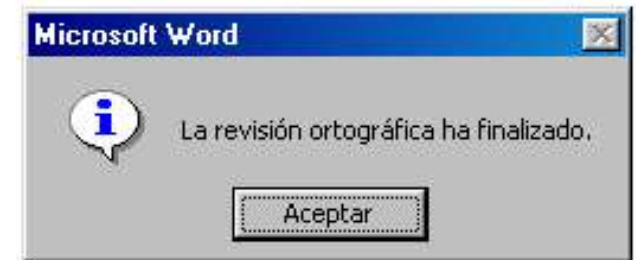
- A.- Ió nou ce s'hi calca li'n formi.
- B.- S'empra pare debate las hales a vans decaure ham pi cat.
- C.- Hi ha pudia cavar d'avenir.
- D.- Bull castiguin qui ets.
- E.- Nom és sec en soc rates nosa viar es.

Comprovem amb  
Word quines  
d'aquestes frases no  
li agraden:



# Ho provem amb Word:

¡Íó nou ce s'hi calca li'n formi.  
S'empra pare debate las hales a vans decaure ham pi cat.  
Hi ha pudia cavar d'avenir.  
Bull castiguin qui ets.  
Nom és sec en soc rates nosa viar es



# Ho provem amb OpenOffice:

Ió nou ce s'hi calca li'n formi.  
S'empra pare debate las hales a vans decaure ham  
pi cat.  
Hi ha pudia cavar d'avenir.  
Bull castiguin qui ets.  
Nom és sec en soc rates nosa viar es



Hem vist que  
l'ordinador no  
pensa. Vegem com  
recorda:

# Necessitem dos voluntaris i una víctima.

Els voluntaris  
proposaran un mot i  
la víctima competirà  
amb l'ordinador;

haurà de decidir si  
el mot és correcte  
segons el diccionari.

**De gran vull ser lingüista**  
23 de maig del 2005  
Universitat Pompeu Fabra

[http://193.145.43.51/dsc/  
Qeimada/askGEC.htm](http://193.145.43.51/dsc/Qeimada/askGEC.htm)

Després de cercar els  
dos mots proposats  
pels voluntaris,  
cerquem un mot  
especial:

***testicprenentelpèlcurtdegambals***



Si tot ha anat bé,  
l'ordinador s'haurà  
mostrat més ràpid  
que l'humà si es  
tracta de mots que  
recorda...

però l'humà és molt  
més ràpid si la  
decisió es basa en el  
coneixement.

*(Aplaudiments per a  
la víctima i els  
voluntaris: plas,  
plas, plas...)*

**De gran vull ser lingüista**  
23 de maig del 2005  
Universitat Pompeu Fabra

*(Prou...)*

Per acabar-ho de  
demostrar, vegem  
quan triga  
l'ordinador a llegir  
tots els noms de lloc  
de Catalunya:

Els ordinadors ens  
poden ajudar a  
escriure millor  
perquè recorden  
quasi instantàniament  
milions de mots.

Però no hi podem  
confiar cegament,  
perquè saben molt  
poques coses dels  
mots que recorden.

La nostra tasca és  
aconseguir que cada  
dia sàpiguen més  
coses sobre les  
paraules, les frases...



# Segona pregunta

Els ordinadors  
saben en quina  
llengua escrivim un  
text?

No podem aplicar  
els avantatges de  
l'ordinador a la  
llengua si no la  
detectem primer.

# Ho comprendrem en conèixer la història del Tiabus, el robot antispam.

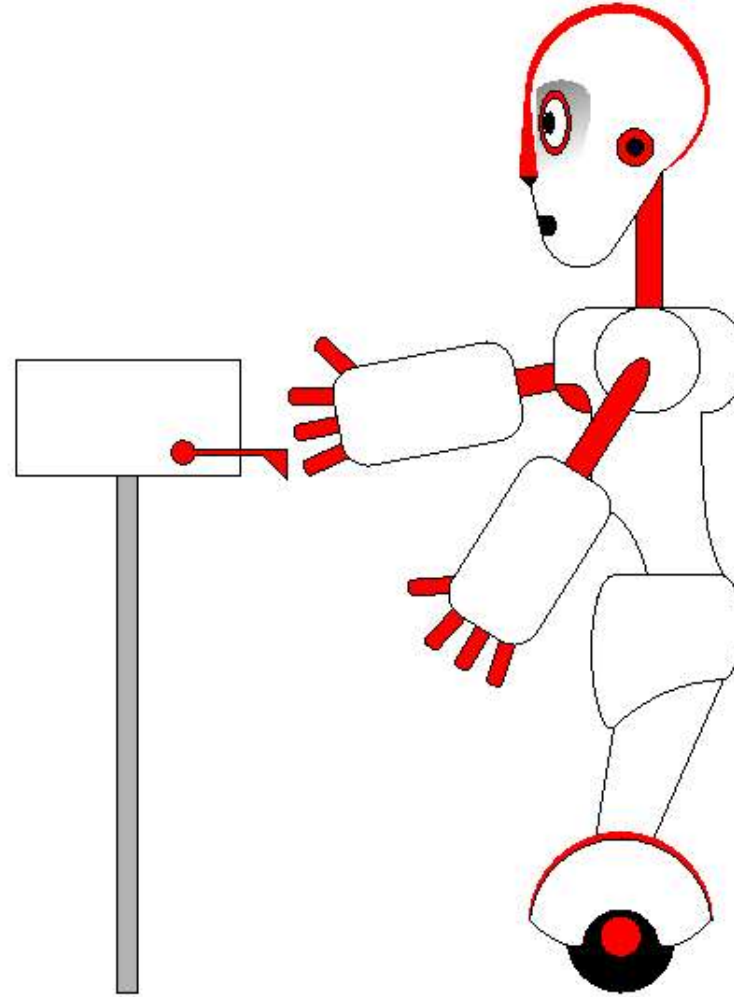
Es tracta d'un robot  
encarregat pel Sr.  
Albert Roig Negre  
per destriar la  
correspondència de  
la publicitat.

El Sr. Albert Roig  
Negre ha encarregat  
la tasca al seu  
eficient secretari  
xinès, el Sr. Li Fa  
Kâs,

que a vegades no  
l'acaba d'entendre  
del tot, però hi posa  
tota la traça.

# El primer dia, en Tiabus agafa el correu del seu amo...

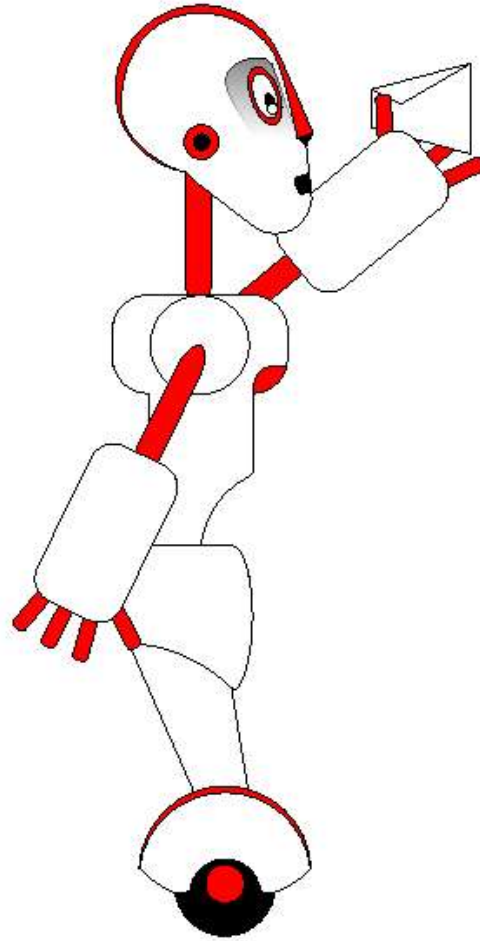
De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra



© dde DEYZA2005

...observa els sobres de les  
cartes...

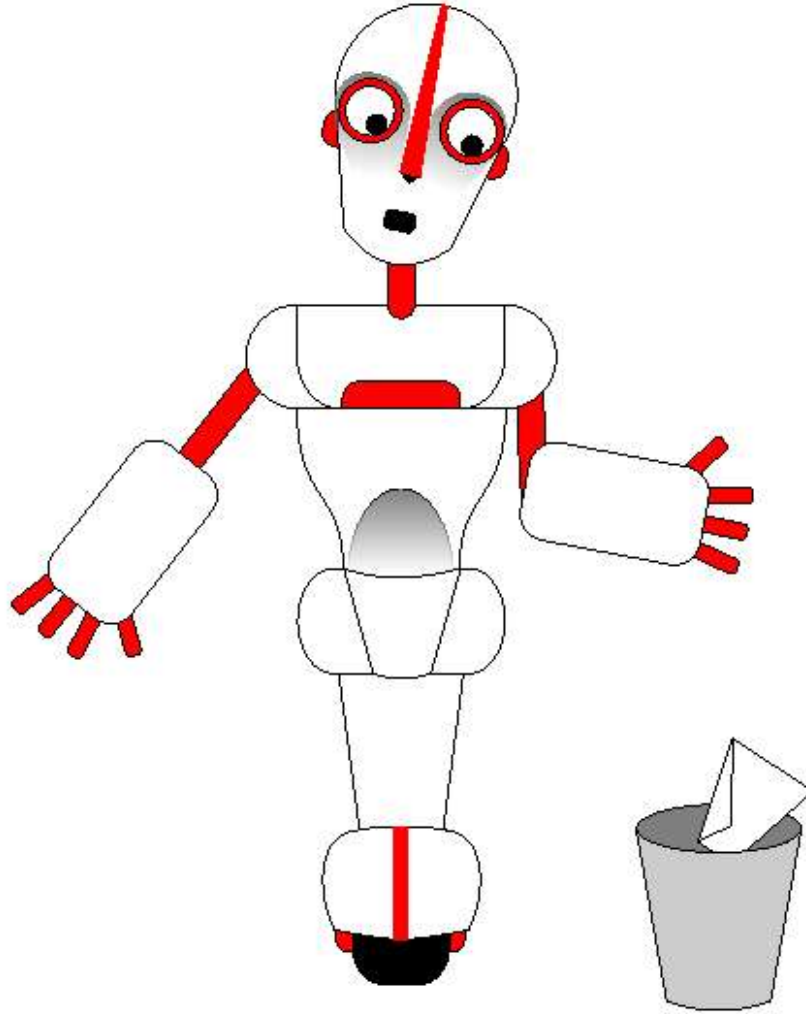
De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra



© ddeDEYZA2005

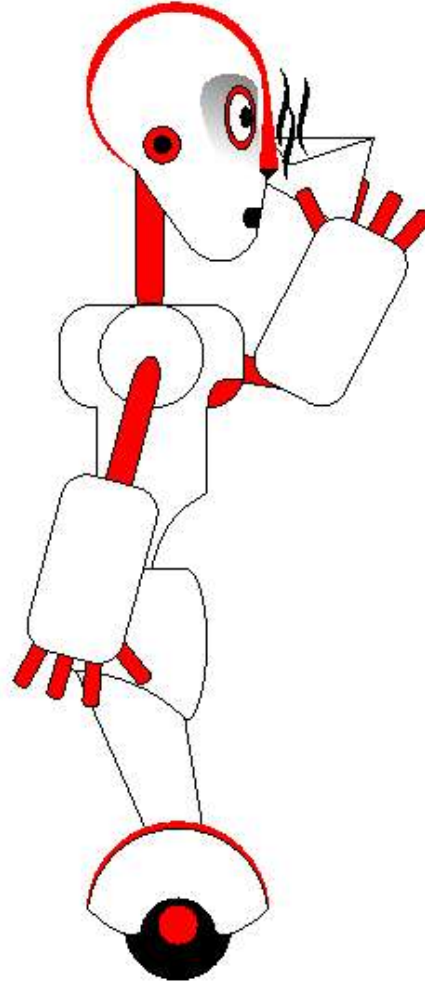


i llença a la paperera els que no  
tenen coloraines: les bones!



© GdeDEYZA2005

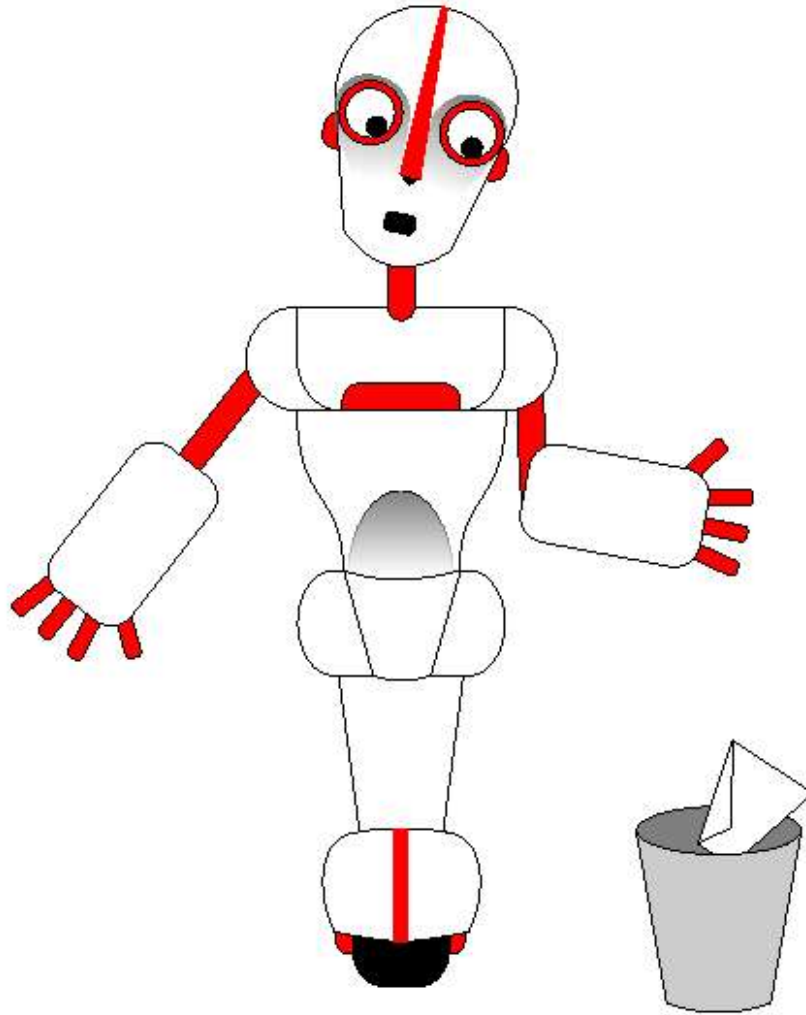
El Sr. Li treballa tota la nit i l'en-  
demà en Tiabus les ensuma...



© ddeDEYZA2005

i llença a la paperera les que  
estan fetes amb paper reciclat!

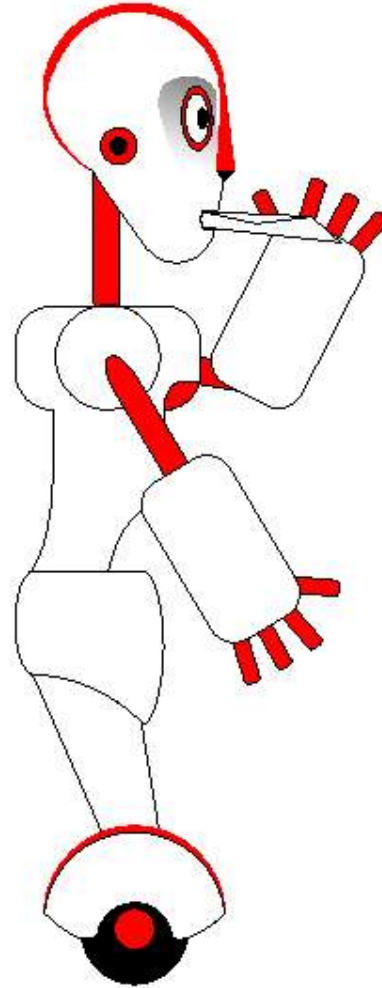
De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra



© GdeDEYZA2005

De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra

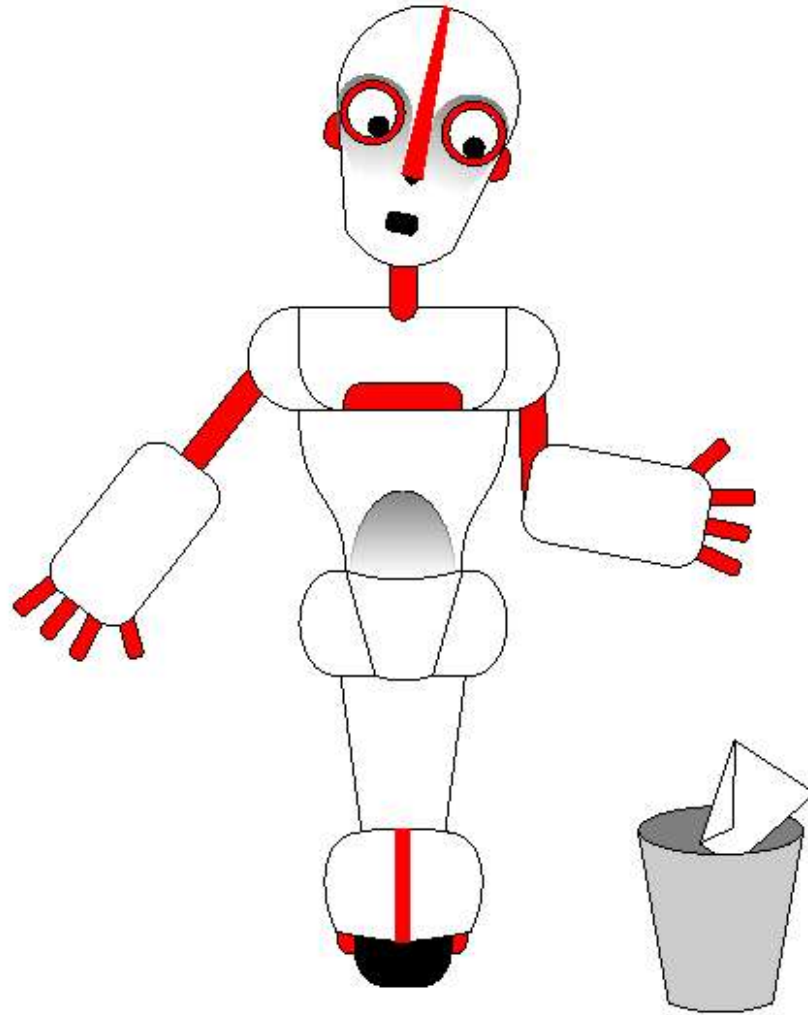
En Li treballa una altra nit i l'en-  
demà en Tiabus tasta les cartes...



© dteDEYZ42005

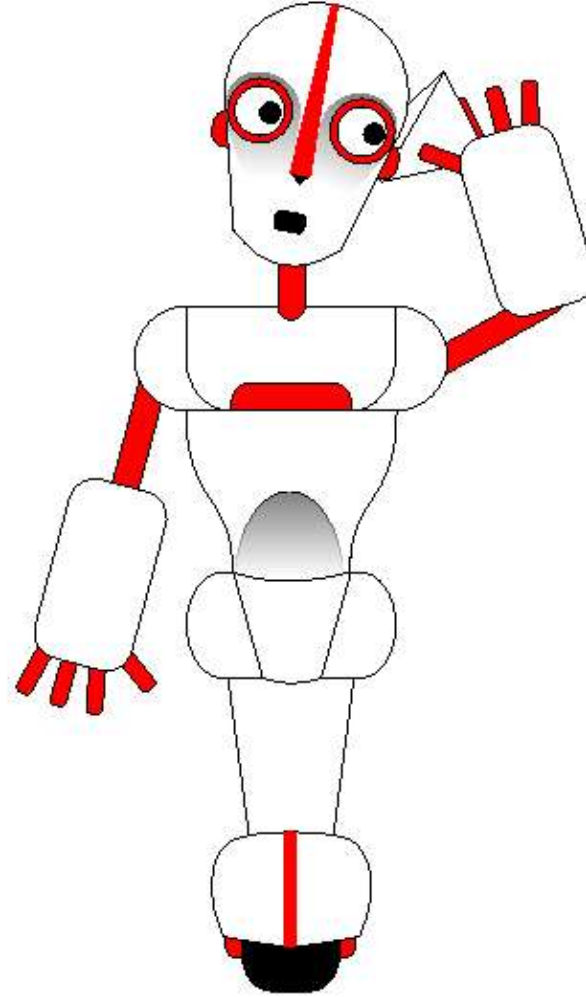
i les llença totes a la paperera  
ben mastegades!

De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra



© GdeDEYZA2005

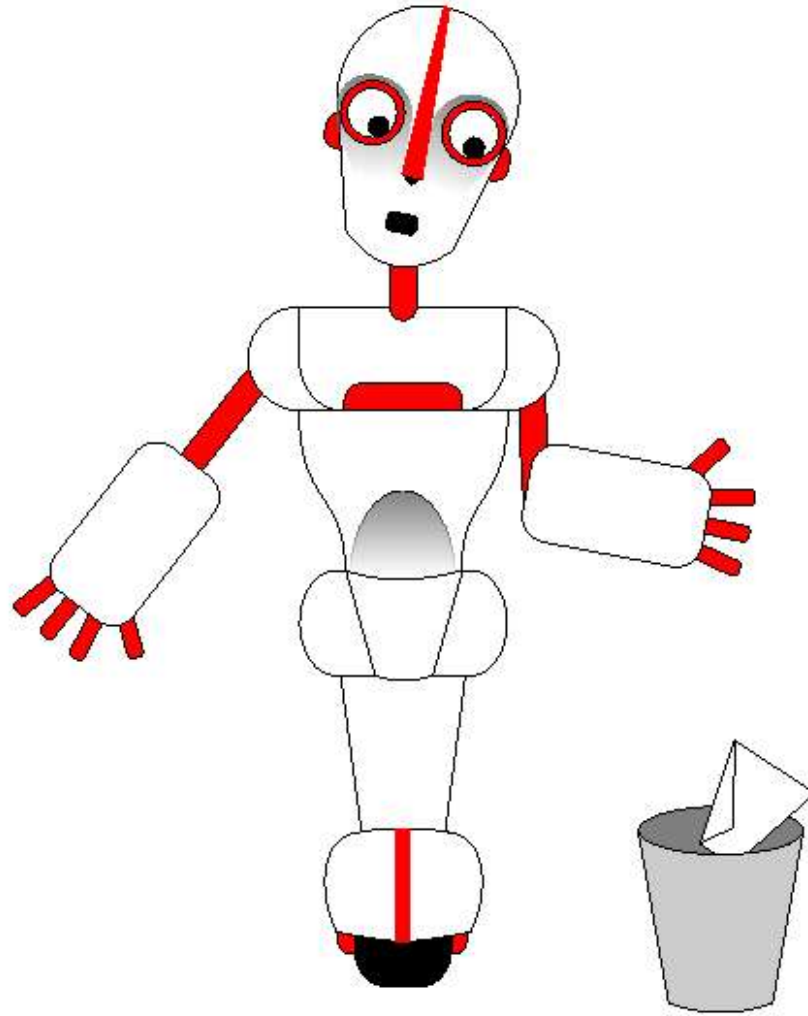
# L'endemà, en Tiabus escolta les cartes...



© Gde DEYZA2005

i només guarda la propaganda  
feta en paper cuixé!

De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra



© GdeDEYZA2005

Arribats aquí, el Sr.  
Li decideix que en  
Tiabus haurà de  
separar les cartes i la  
publicitat llegint-les.



Per això estem elaborant un detector de llengües, que també servirà per filtrar els correus electrònics.

# Li donem un text:

Declaració Universal de Drets Humans PREÀMBUL Considerant que el reconeixement de la dignitat inherent i dels drets iguals i inalienables de tots els membres de la família humana és el fonament de la llibertat, la justícia i la pau en el món, Considerant que el desconeixement i el menyspreu dels drets humans han originat actes de barbàrie que han ultratjat la consciència de la humanitat; i que s'ha proclamat com l'aspiració més elevada de tothom l'adveniment d'un món on els éssers humans, deslliurats del temor i la misèria, puguin gaudir de llibertat d'expressió i de creença, Considerant que és essencial que els drets humans siguin protegits per un règim de dret per tal que les persones no es vegin forçades, com a últim recurs, a la rebel·lió contra la tirania i l'opressió, Considerant també que és essencial de promoure el desenvolupament de relacions amistoses entre les nacions, Considerant que els pobles de les Nacions Unides han ratificat en la Carta llur fe en els drets humans fonamentals, en la dignitat i el valor de la persona humana i en la igualtat de dret d'homes i dones; i que han decidit de promoure el progrés social i millorar el nivell de vida dins d'una llibertat més àmplia, Considerant que els Estats membres s'han compromès a assegurar, en cooperació amb l'Organització de les Nacions Unides, el respecte universal i efectiu dels drets humans i les llibertats fonamentals, Considerant que una concepció comuna d'aquests drets i llibertats és de la més gran importància per al ple compliment d'aquest compromís, L'Assemblea General Proclama aquesta Declaració Universal de Drets Humans com l'ideal comú a assolir per a tots els pobles i nacions amb el fi que cada persona i cada institució, inspirant-se constantment en aquesta Declaració, promoguin, mitjançant l'ensenyament i l'educació, el respecte a aquests drets i llibertats i assegurin, amb mesures progressives nacionals i internacionals, el seu reconeixement i aplicació universals i efectius, tant entre els pobles dels Estats membres com entre els dels territoris sota llur jurisdicció.

# triem algunes opcions...

De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra

**LATEL** Stop Words

Write or copy a sentence in any language:

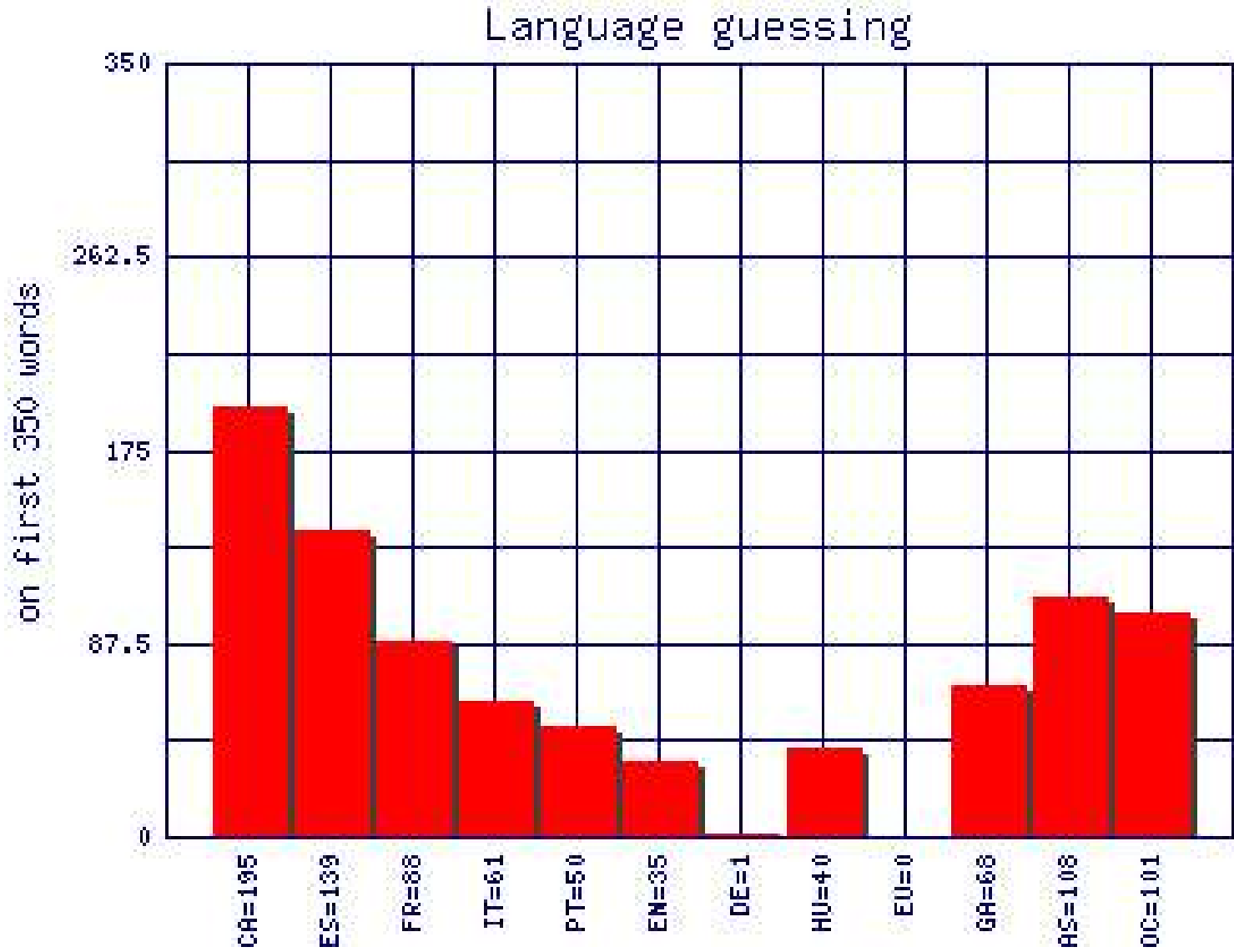
Declaració Universal de Drets Humans PREÀMBUL  
Considerant que el reconeixement de la dignitat inherent i dels drets iguals i inalienables de tots els membres de la família humana és el fonament de

Checking: full>=500 Bars

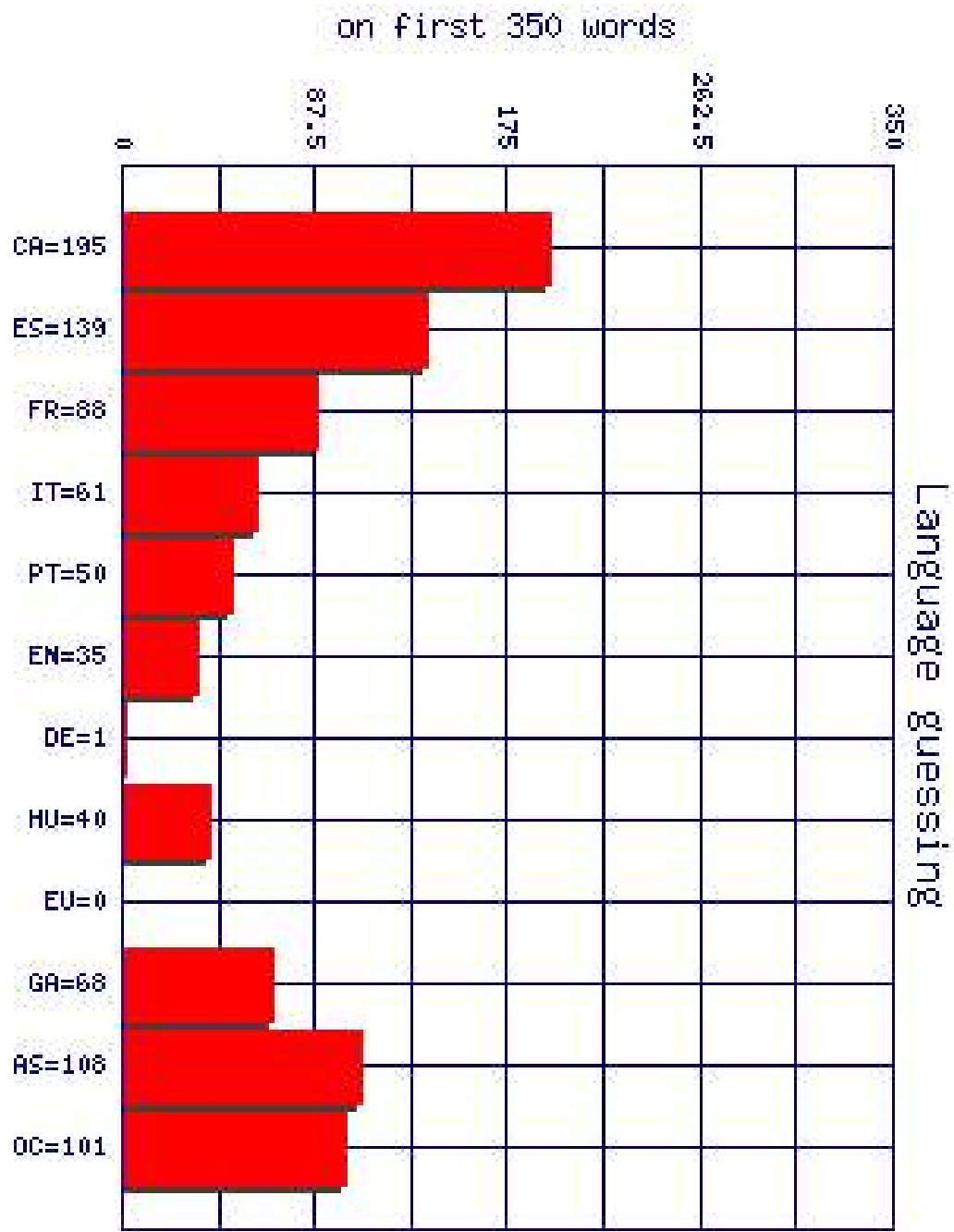
Guess!



i el detector ens diu  
quantes paraules  
reconeix de cada  
llengua:



De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra



Ho fa cercant les  
paraules més  
freqüents de cada  
llengua.

I com podem saber  
quines són les  
paraules més  
freqüents del català?



# Els nostres col·legues han introduït a l'ordinador textos representatius

fins a un total de  
60 milions de  
paraules. N'hem  
extret les més  
freqüents.

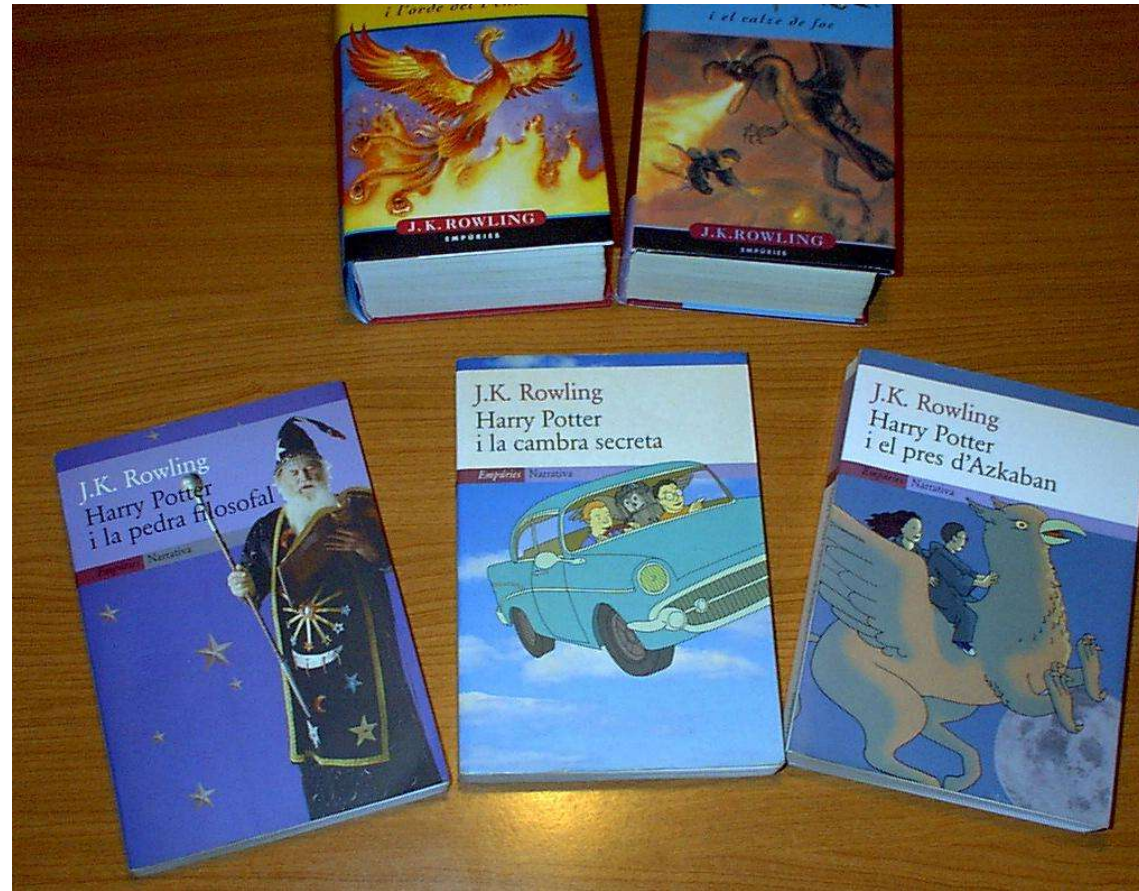
60.000.000 de  
mots a 3.000 x full  
= 20.000 fulls: la  
pila de caixes !!!

En comparació, els  
cinc llibres de  
Harry Potter tenen  
més o menys  
1.221.480 mots.

$(968+696+271+301+374)*13*36=1.221.480$

# Quant triguem a llegir-los?

De gran vull ser lingüista  
23 de maig del 2005  
Universitat Pompeu Fabra



L'ordinador llegeix  
60 milions de mots  
en menys d'una hora.

Els recorda tots,  
però no n'entén cap.

# Ens ajuda a trobar els mots més freqüents de cada llengua



I comprovant a cada correu els mots més freqüents, en Tiabus podrà saber amb quin diccionari l'ha d'interpretar.

En resum: els  
ordinadors tenen  
més memòria que  
intel·ligència; són  
lloros molt ràpids.

Però cada vegada  
en saben més i cada  
vegada ens ajuden  
més.

**De gran vull ser lingüista**  
23 de maig del 2005  
Universitat Pompeu Fabra

**Fi**

Post Scriptum:  
als lingüistes la  
llengua ens agrada  
fins per jugar-hi...

La torna:  
Albert Roig Negre  
Tiabus  
Li Fa Khâs