

# LEGOLAS 2.0 COM A EINA D'EXTRACCIÓ DE SEQÜÈNCIES D'ETIQUETES MORFO-SINTÀCTIQUES PER A LA COMPARACIÓ FORENSE DE TEXTOS ESCRITS



M. Teresa Turell i Maria S. Spassova  
ForensicLab  
Institut Universitari de Lingüística Aplicada  
Universitat Pompeu Fabra  
iulaforensic@upf.edu  
<http://www.iula.upf.edu/forensiclab/fpresca.htm>



## APLICACIONS DE L'EINA

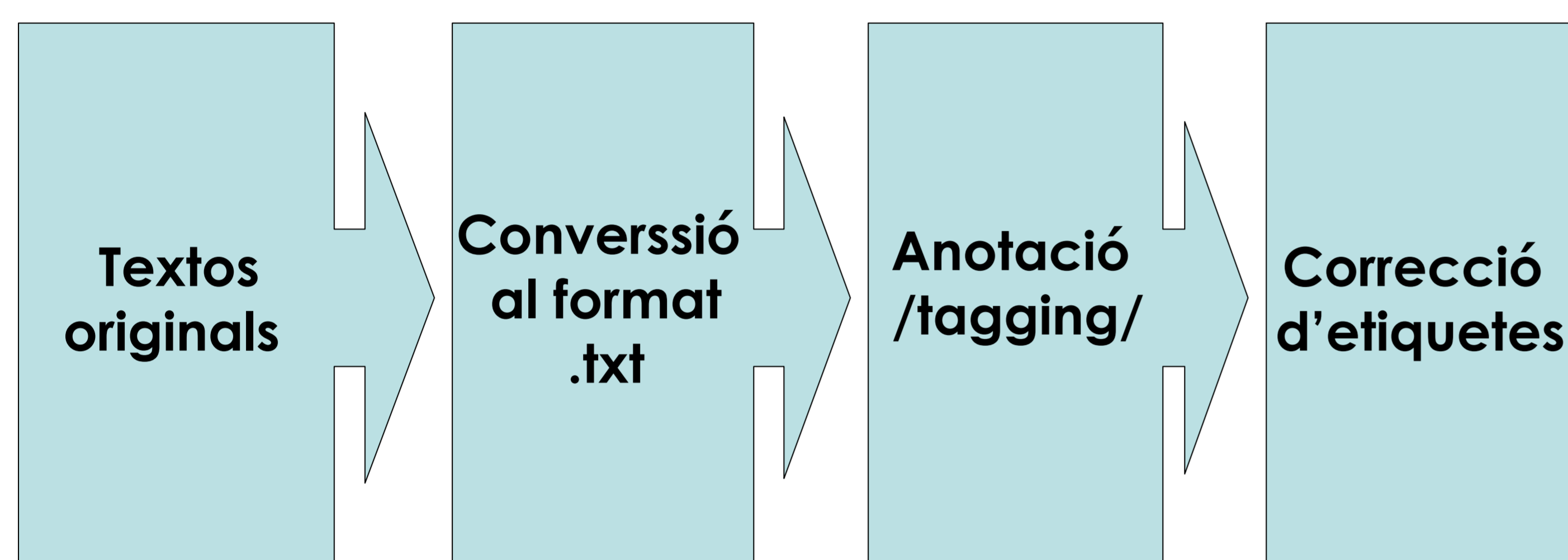
- Comparació forense de textos escrits útil per a una possible atribució/determinació d'autoria
- Experimentació dins l'àmbit de l'idiolectometria amb aplicacions forenses
- Classificació de textos

## Captura de pantalla del fitxer de resultats extrems amb l'eina

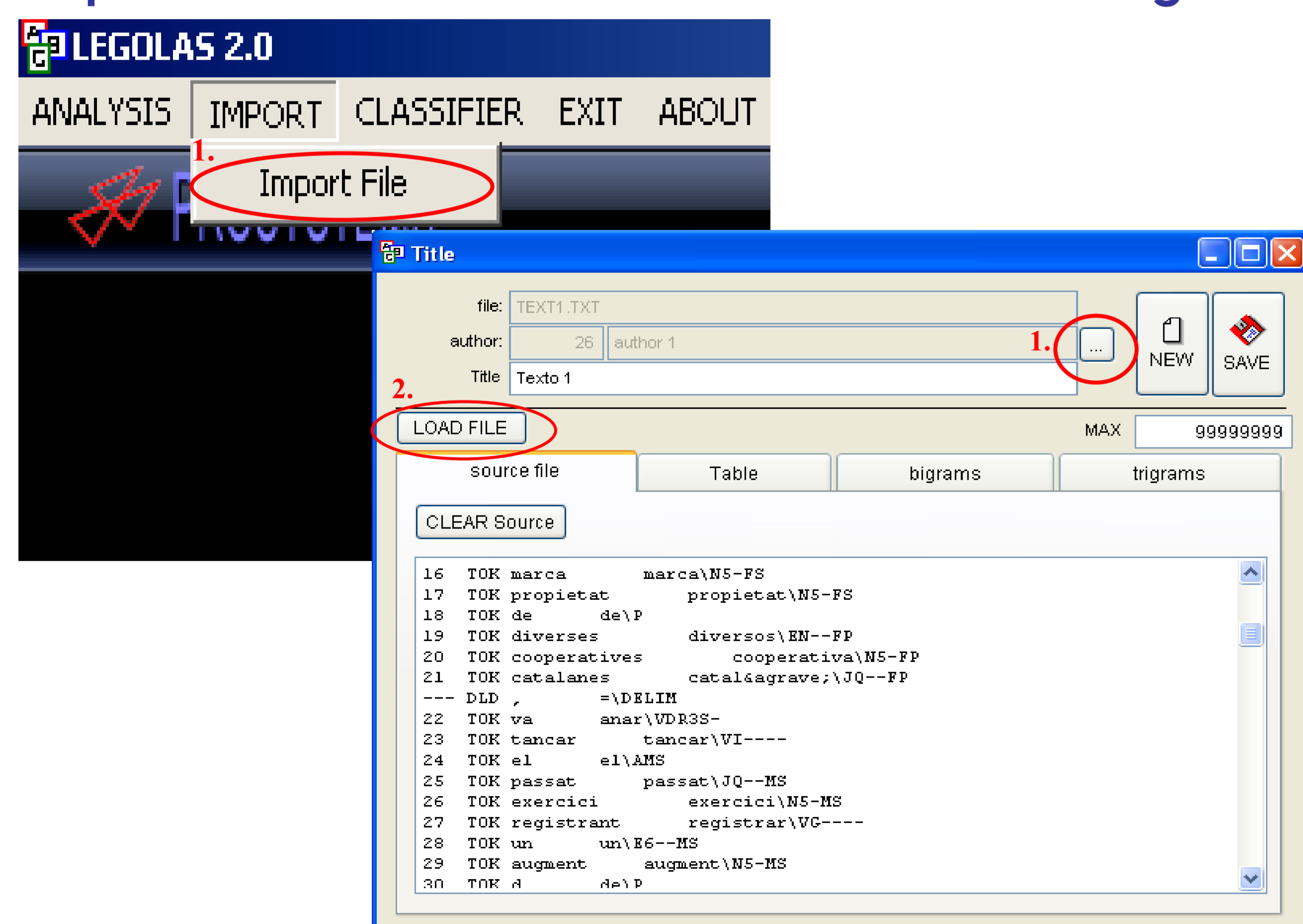
Autor	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Autor 3															
2. bigrams	texto 1	texto 2	texto 3	texto 4	texto 5	texto 6	texto 7	texto 8	texto 9	texto 10	texto 11	texto 12	texto 13	texto 14	texto 15
3. NP	69	60	30	43	25	26	51	65	53	26	43	52	27	39	
4. P.NG	23	39	32	37	27	30	41	29	43	34	36	32	16	41	
5. AMS.NG	36	36	26	37	24	24	21	51	27	22	26	27	9	21	
6. NE.JQ	7	15	22	31	18	19	21	25	23	20	24	29	25	32	
7. P.AMS	32	39	20	25	32	17	19	36	20	14	26	24	6	17	
8. AFS.NG	34	21	20	24	8	13	28	21	24	11	28	26	16	14	
9. P.AFS	13	15	21	25	9	18	19	17	24	16	21	22	10	17	
10. NE.C	23	15	18	18	22	25	13	20	16	9	16	9	5	12	
11. P.VA	9	16	13	10	11	10	8	13	12	8	15	12	12	11	
12. JQ.NG	19	7	12	15	17	20	10	14	10	11	14	7	7	8	
13. JQ.P	1	10	10	6	9	7	11	11	6	11	22	19	10	14	
14. D.P	5	11	9	17	5	5	13	11	20	7	12	12	16	13	
15. P.ISS.NG	7	10	12	12	7	5	11	12	14	3	16	14	11	13	
16. NE.D	6	11	9	14	9	7	7	11	15	10	16	17	10	5	
17. C.D	2	8	3	7	9	7	10	3	9	5	4	20	4	7	
18. A.NG	4	5	11	6	8	6	11	11	9	13	5	3	7	11	
19. V.P	4	5	2	7	9	3	9	7	13	6	9	12	7	14	
20. NE.RR	3	13	4	5	4	4	5	11	6	8	13	5	3	4	
21. ES.NG	6	13	5	5	6	6	4	7	5	9	4	12	10	8	
22. C.P	2	4	2	10	6	0	5	9	5	7	9	8	4	10	
23. P.IPES	4	8	2	4	7	4	8	5	7	6	2	12	4	11	
24. AMP.NG	4	3	6	4	7	9	8	1	5	2	7	4	8	4	
25. D.VORS	2	4	6	12	4	5	9	6	5	6	7	8	1	7	
26. V.P	4	12	2	14	5	3	2	3	8	11	10	8	5	4	
27. P.APP	0	5	5	5	5	2	3	4	6	5	4	6	10	11	
28. P.ISS.NG	1	7	7	7	3	4	4	3	4	2	9	6	8	8	
29. AFP.NG	8	4	6	5	5	5	7	0	6	4	7	4	5	1	
30. P.REU	3	4	1	11	1	4	8	4	6	2	5	3	6	3	
31. EN.NG	4	5	5	3	2	4	5	6	1	5	11	8	2	4	
32. P.IPES	3	6	3	0	7	1	5	3	5	3	4	5	3	5	
33. VORS.P	0	5	5	1	3	6	7	3	9	5	4	5	4	5	
34. JQ.C	3	6	7	2	4	2	8	2	5	2	6	2	4	5	
35. P.AMP	2	4	5	5	3	7	1	3	3	0	1	7	10	6	
36. P.EN	1	2	2	6	8	4	7	0	4	4	5	1	0	1	
37. C.NG	2	2	5	1	1	2	5	6	3	4	5	1	2	3	
38. C.VI	2	0	3	2	5	7	0	4	1	2	5	3	4	3	
39. D.D	0	2	6	8	2	3	3	3	4	4	1	1	3	1	
40. P.ISS.NG	2	2	4	9	1	3	0	3	2	5	2	1	0	1	

## PRESTACIONS PER AL TRACTAMENT COMPUTACIONAL DE TEXTOS ANOTATS

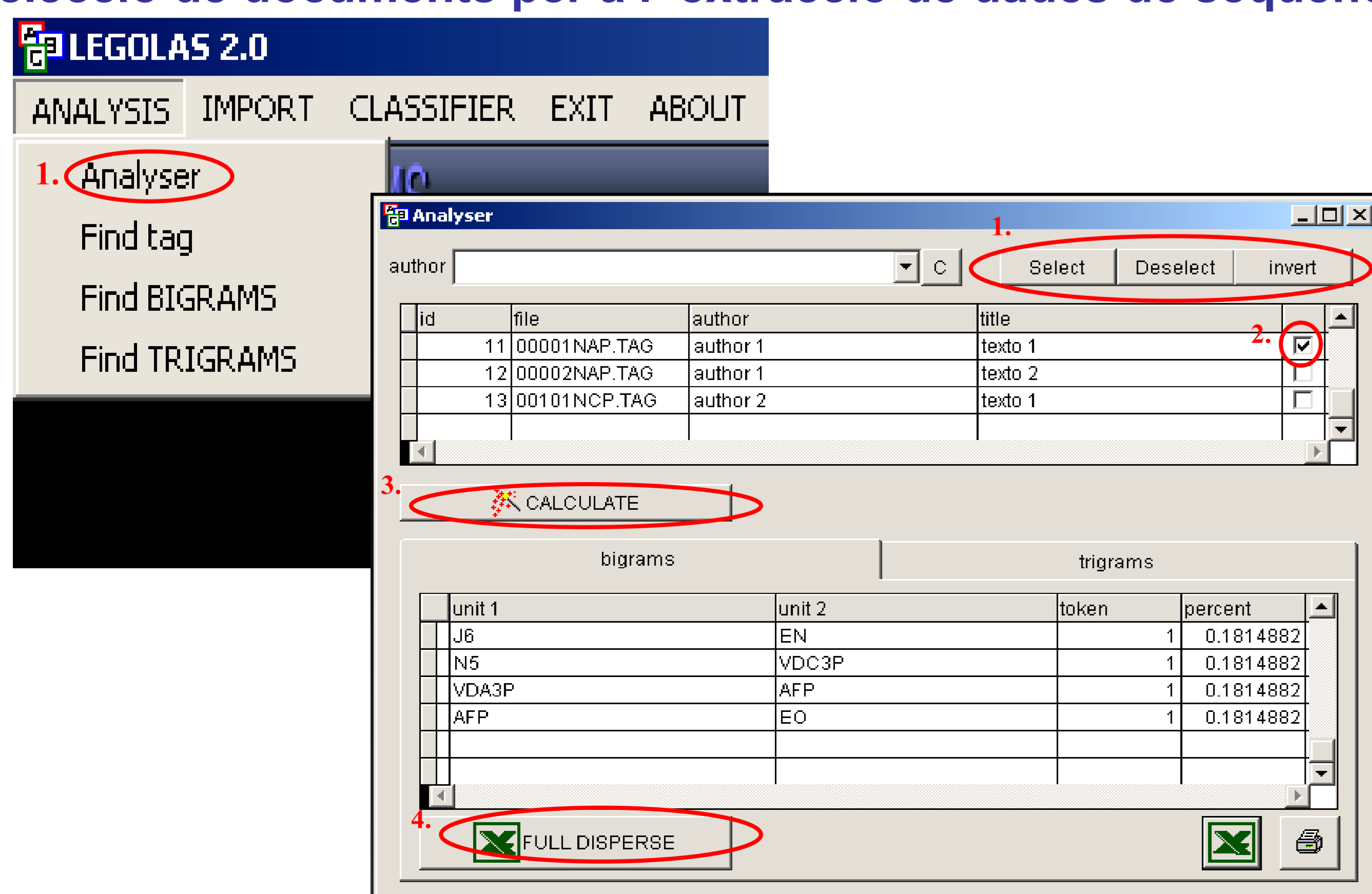
### Esquema del tractament previ a l'extracció de dades



### Incorporació de textos a la base de dades de Legolas 2.0



### Selecció de documents per a l'extracció de dades de seqüències

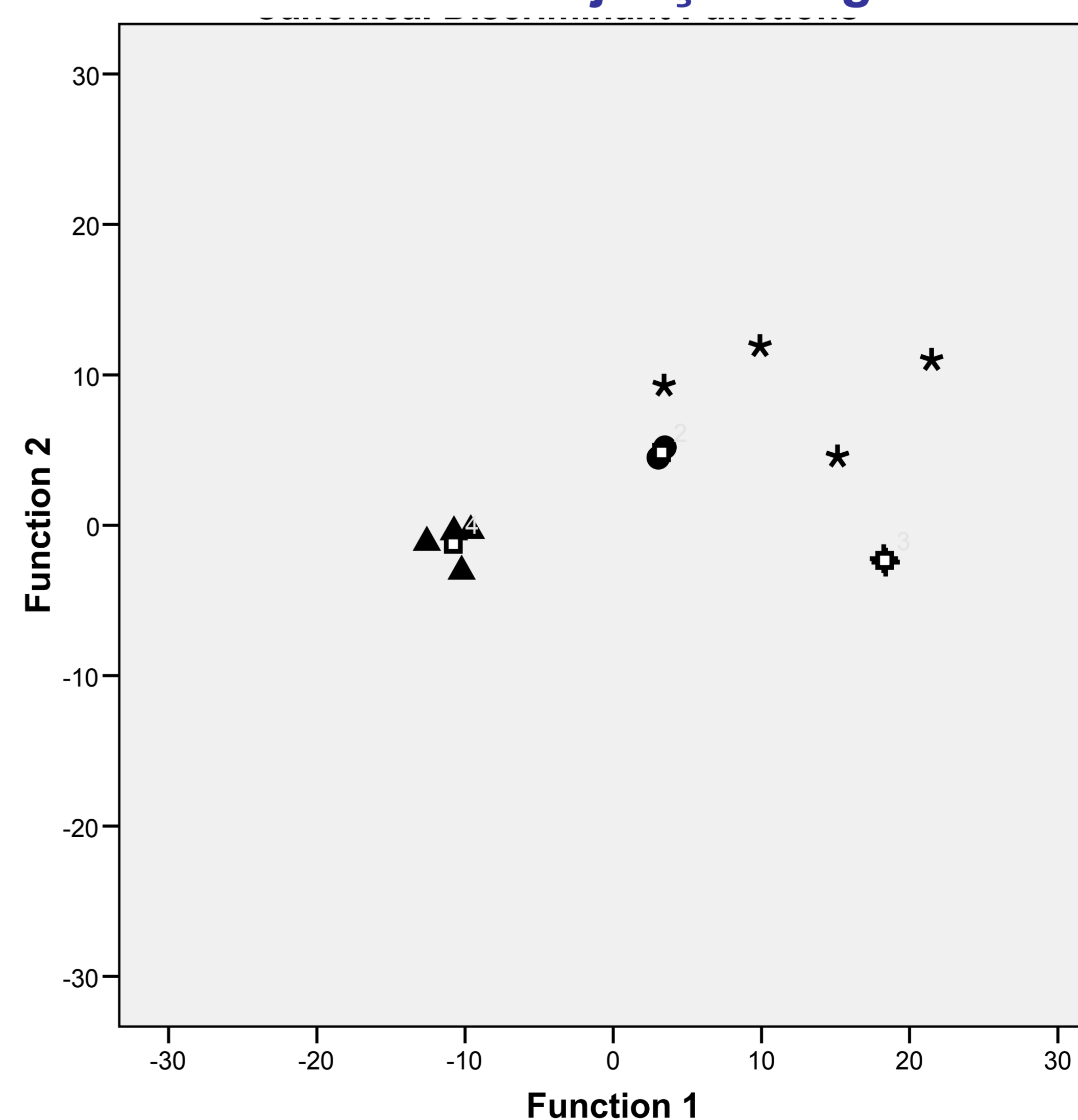


## RESULTATS DE LA CONSULTA DE CORRESPONDÈNCIA 'SEQÜÈNCIA - UNITAT LINGÜÍSTICA'

unit 1	word 1	unit 2	word 2	unit 3	word 3
P	a , a	AMS	i , et	N5	referent , refe
P	a , a	AMS	i , et	N5	mercat , merc
P	de , de	AMS	i , et	N5	sector , secto
P	a , a	AMS	i , et	N5	mercat , merc
P	de , de	AMS	i , et	N5	pa&iacute;s , j
P	de , de	AMS	i , et	N5	preu , preu
P	de , de	AMS	i , et	N5	mercat , merc
P	de , de	AMS	i , et	N5	territori , territ
P	amb , amb	AMS	el , el	N5	territori , territ
P	de , de	AMS	i , et	N5	dret , dret
P	de , de	AMS	i , et	N5	dret , dret
P	en , en	AMS	el , el	N5	decurs , decu
P	de , de	AMS	i , et	N5	temps , temps
P	de , de	AMS	i , et	N5	projecte , proj

## PROJECCIÓ DELS RESULTATS DE L'APLICACIÓ DE LEGOLAS 2.0 A LA COMPARACIÓ FORENSE DE TEXTOS

### Classificació mitjançant bigrames\*



- Llegenda: \* textos dubitats  
● textos indubitats  
▲ textos de control