

CAPÍTOL 10

LLARGADA DE CAPÍTOL

Índex Capítol

10.1	Descripció de les dades.....	143
10.2	Estimació del punt de canvi.....	147
10.3	Conclusió.....	148

CAPÍTOL 10

LLARGADA DE CAPÍTOL

Coromines (1956) assegura, parlant del possible paper jugat per Galba, que

“estem sens dubte davant el cas de l’arreglador indiscret, encarregat de la preparació per a la impremta d’un original pràcticament llest, però que de sobte concep l’ambició de convertir-se en coautor d’un llibre famós, i per tal d’inflar la seva part no vacil·la a servir-se de tota mena de materials, a recórrer a l’argúcia de tallar més curts els capítols de la seva part, perquè vegi l’impressor que ha escrit més de la tercera part del que té el llibre: als 170 capítols de la part de Galba corresponen 344 pp. de l’ed. Riquer (=2 pp. per capítol), als 320 de Martorell corresponen 900 pp. (gairebé 3 per capítol)”.

Per tant, sembla que l’estudi de la llargada de capítol pugui indicar una possible frontera d’estil en el *Tirant*. Des del punt de vista estadístic no tenim constància de que s’hagi utilitzat la llargada de capítol com a mesura per a determinar l’autoria de texts.

10.1 Descripció de les dades

Mesurem la llargada de capítol, N_i , en nombre de paraules. Tal i com ja s’ha esmentat en el capítol 4, no s’inclouen ni els títols ni aquelles paraules que en la edició de Riquer(1983) apareixen en cursiva.

L’histograma de la Figura 10.1 mostra la distribució de les llargades de capítol en el *Tirant*. La forma de la distribució és clarament no simètrica, amb una cua molt llarga i el màxim es troba per llargades entre 300 i 400 paraules. De fet, un capítol té acotada la seva llargada mínima en 0 paraules, però no la seva llargada màxima, i això es reflexa en l’histograma. També es pot veure que hi ha capítols, dinou, de llargada 0, i altres

d'extraordinàriament curts, 24 capítols de menys de 100 paraules. Els capítols de llargada nul·la són aquells en els que totes les paraules apareixen en cursiva.

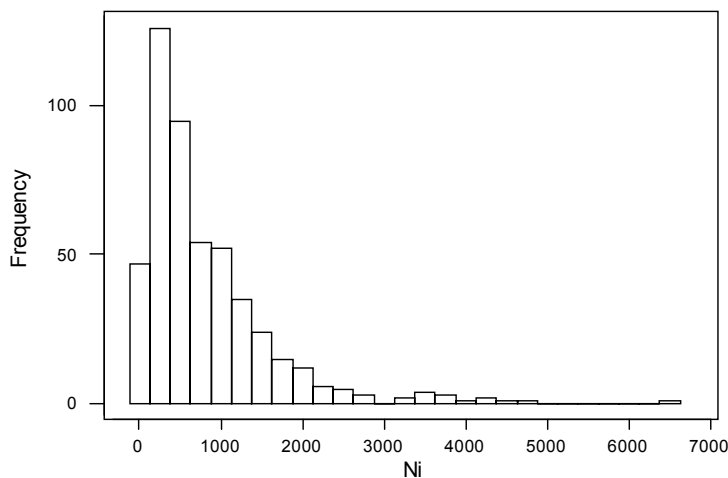


Figura 10.1: Histograma de les llargades dels capítols del Tirant, N_i , mesurades en nombre de paraules.

Representem en la figura 10.2 el diagrama de punts de les llargades de capítol, estratificat en dos subgrups, el primer format pels primers 324 capítols i el segon pels capítols des del 325 al final, basant-nos en les paraules de Coromines com a indicació d'on pot estar la possible frontera d'estil. De l'anàlisi gràfica dels diagrames de punts observem com:

- El màxim de les distribucions és força coincident en els dos casos,
- pels capítols anteriors al 325 la cua és molt més llarga, mentre que pels posteriors la llargada pràcticament s'atura en 2000 paraules,
- La proporció de capítols molt curts és més gran pels capítols de 1 a 325,
- La distribució és més apuntada pels capítols posteriors al 325.

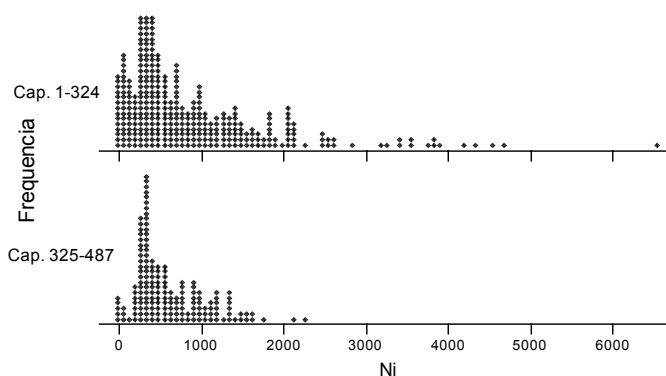


Figura 10.2: Diagrama de punts de les llargades dels capítols del Tirant estratificat en funció de que es trobin abans o després del capítol 325.

L'evolució temporal de la llargada del capítol, figura 10.3, ens mostra com els capítols més llargs de 3000 paraules es troben entre el 68 i el 164 (bàsicament entre el 100 i el 164), amb dues excepcions, el 189, que és el més llarg del llibre amb 6521 paraules, i el 262 amb 3426 paraules. Trobem només dos capítols posteriors al 325 de més de 2000 paraules, el 340 i el 349, mentre que n'hi ha 22 d'anteriors

La llargada mitjana de capítol, calculada considerant els capítols amb llargada no nul·la, és de 847 paraules/cap. 125 capítols anteriors i 45 posteriors al 325 superen la mitjana,. Observem com les proporcions de capítols més llargs que la mitjana són semblants pels capítols que es troben abans i després del 325.

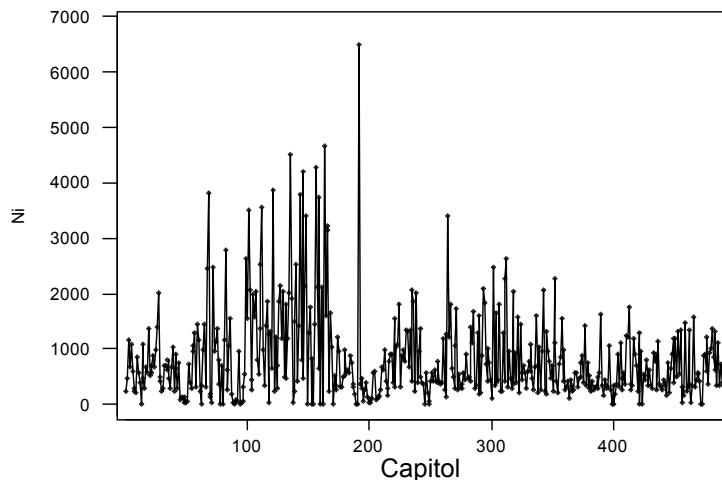


Figura 10.3: Evolució de la llargada de capítol, N_i . Hi ha representats tots els capítols
 La distribució del logaritme de N_i és més simètrica que la de N_i , i les cues es redueixen, tal i com mostra l’histograma de la figura 10.4.

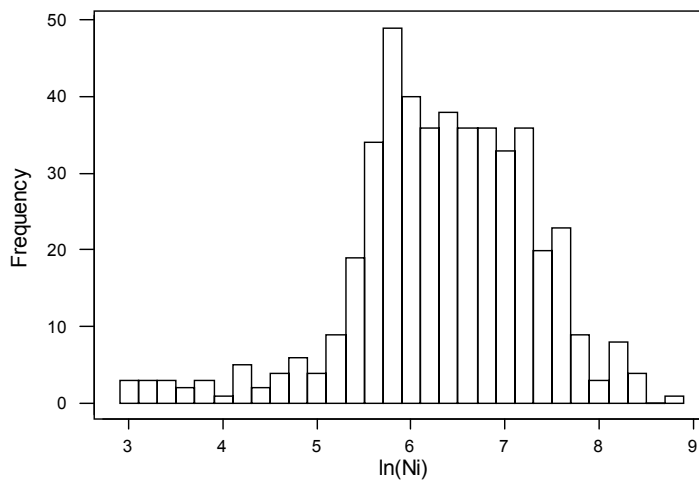


Figura 10.4: Histograma del logaritmes de les llargades dels capítols amb $N_i > 0$ del Tirant.

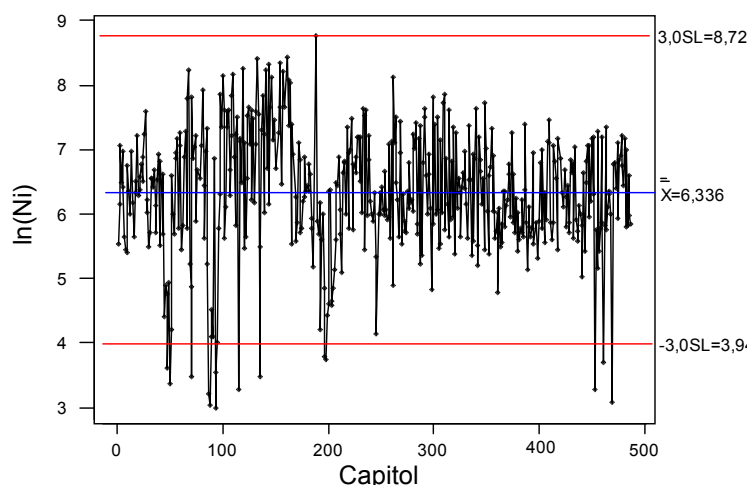


Figura 10.5: Gràfic de Control per Observacions Individuals per la llargada de capítol mesurada com el logaritme del nombre de paraules $\log(N_i)$. Hi ha representats tots els capítols amb $N_i > 0$.

El gràfic de control per observacions individuals de la figura 10.5 mostra l'evolució temporal del logaritme de la llargada dels capítols. En ell es veu una tendència a créixer, entre l'inici i el capítol 164, junt amb una presència notable de capítols molt curts: hi ha 15 capítols de menys de 100 paraules a més a més de 9 de llargada nul·la anteriors al capítol 164). A partir d'aquest punt, s'observa una tendència a disminuir que s'allarga fins al capítol 198, i que només es trencada per la presència del capítol més llarg de tot el *Tirant*, el 189. Continúa amb una tendència a augmentar fins al capítol 236 per establitzar-se. En els capítols finals, a partir del 450, es torna a trobar una concentració elevada de capítols extraordinàriament curts: n'hi ha 3 de menys de 50 paraules. Entre els capítols 250 i 450 no n'hi ha cap amb menys de 100 paraules, tret de 4 amb llargada nul·la, i només 7 amb llargada entre 100 i 200 paraules.

La Figura 10.6 mostra el gràfic CUSUM pel logaritme de la llargada de capítol, per $N_i > 0$. En el capítol 5 s'ha descrit com per cada capítol m es calcula l'estadístic:

$$S_m = \sum_{i=1}^m (\ln(N_i) - \mu_0)$$

on $\ln(N_i)$ és el logaritme de la llargada del capítol, S_m rep el nom de suma acumulada fins al capítol m i μ_0 és el valor nominal o objectiu. Com que no hi ha un valor objectiu, s'ha triat com a μ_0 la mitjana dels logaritmes de les llargades de capítol.

El gràfic mostra una clara tendència de la suma acumulada a decreixer fins al capítol 97 pel que la llargada d'aquests capítols és, en general, més petita que la llargada mitja de capítol, després augmenta fins al capítol 167, el que indica que la llargada dels capítols és, en general, més gran que la llargada mitja. Entre el 167 i el 189 es el gràfic és força pla, pel que la llargada d'aquests capítols és semblant a la llargada mitja dels capítols. El màxim del gràfic es té pel capítol 189, que és el més llarg del *Tirant*. a partir d'aquí el gràfic passa per un mínim per tornar a créixer fins al capítol 357, on s'obté un màxim local amb valor de la suma acumulada semblant al del màxim global. A partir d'aquest capítol la suma acumulada decreix fins a valer 0 per al darrer capítol.

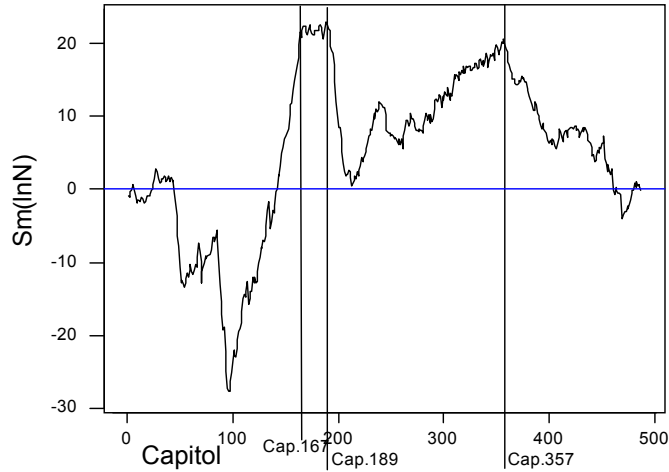


Figura 10.6: Gràfic CUSUM pel logaritme de la llargada de capítol, $\ln(N_i)$. S’ha pres com a valor “objectiu” la mitjana dels capítols amb $N_i > 0$. El màxim de la suma acumulada es té pel capítol 189.

S’obté un gràfic molt semblant si es representa la suma acumulada de la llargada de capítol en lloc del seu logaritme.

10.2 Estimació del punt de canvi

Per a l’estimació del punt de canvi en la seqüència dels logaritmes de les llargades de capítol per N_i no nul·les, s’han fet servir les tècniques descrites al capítol 6 per a una seqüència de normals, i que ja han sigut utilitzades en el cas de la llargada de paraula i de frase.

Assumint independència entre les observacions, s’ha ajusten els $n-1$ models:

$$\ln(N_i) \sim N(\beta_0^{(r)} + \beta_1^{(r)} \text{Ind}_{1i}^{(r)}, \sigma)$$

on $\text{Ind}_{1i}^{(r)}$ és una variable indicadora que pren valor 0 per $i=1,2,\dots,r$ i pren valor 1 per $i=r+1,\dots,n$. Estimem el punt de canvi com:

$$\hat{r}_N = \max_{1 \leq r \leq n-1} F_r$$

on F_r és el valor de l’estadístic F de la taula ANOVA pel model amb variable indicadora $\text{Ind}_{1i}^{(r)}$

El gràfic de la figura 10.7 mostra l’evolució de F_r en funció de r . S’observa, bàsicament, el mateix que es veia en el gràfic CUSUM: el màxim es té per a $r=97$, després el valor cau en picat, torna a augmentar prenent valors molt semblants entre els capítols 167 i 189, a continuació decreix, per augmentar fins al 355, a partir d’on torna a decreixer. El valor màxim de F_r és 11,05, que equival a un valor de $t=3,3$ per a l’estadístic per contrastar la hipòtesi que no hi ha diferències entre les mitjanes dels logaritmes de les llargades dels capítols anteriors i posteriors a $r=97$. Aquest valor no ens assegura que la diferència de mitjanes sigui significativa.

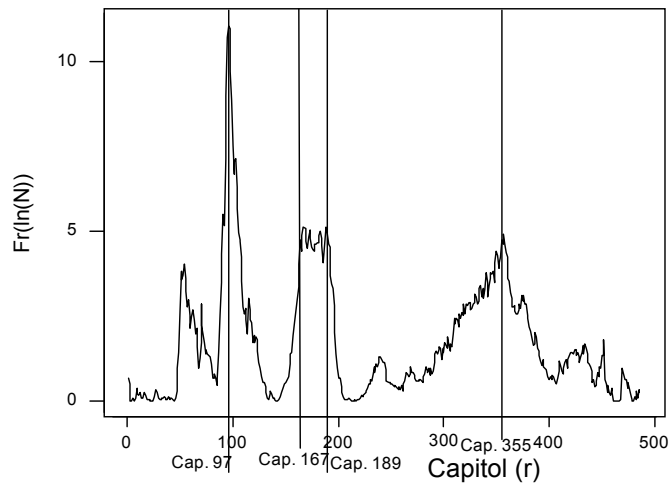


Figura 10.7: Gràfic de l'evolució de F_r en funció de r per a $\ln(N_i)$.

Resultats molt semblants s'obtenen si s'estima pel mateix mètode el punt de canvi per a la seqüència univariant de llargades de capítol en lloc del seu logaritme.

10.3 Conclusió

Tot i que en l'anàlisi gràfica s'observen diferències en la distribució de la llargada de capítol entre els capítols del principi i els del final del llibre, es fa difícil assegurar que aquestes diferències siguin significatives. A més a més no es demostra l'existència d'un punt en el que la distribució de la llargada de capítol canvi.

No es possible, mitjançant l'estudi estilomètric de la llargada de capítol, demostrar la doble autoria del *Tirant*, tal i com assegura Coromines (1956).