

# CAPÍTOL 14

## CONCLUSIONS

### Índex Capítol

14.1 Punt de canvi per capítols .....	259
14.2 Assignació de capítols a un dels dos estils .....	260
14.2.1 Resum dels resultats de les anàlisi clusters individuals.....	261
14.2.2 Anàlisi clusters individuals imposant grandària 100 .....	264
14.2.3 Cluster amb llargada de paraula, ús de paraules i diversitat.....	265
14.3 Conclusions .....	266
Annex 14.1: Resum de les assignacions de capítols a clusters.....	267
Annex 14.2: Assignacions de capítols a clusters fixant la grandària dels clusters.....	277
Annex 14.3: Assignació de capítols mitjançant una anàlisi a clusters per variables contínues.....	287



# CAPÍTOL 14

## CONCLUSIONS

En els capítols del 8 al 13 hem fet servir les tècniques, basades en l'ajust de models i explicades en el capítol 6, per estimar el capítol on hi ha el punt de canvi,  $\hat{r}$ , en les seqüències de normals, binomials i multinomials formades pels valors que prenen les unitats d'estadística textual analitzades al llarg de *Tirant*. A través de l'anàlisi cluster, s'ha observat com capítols i blocs posteriors al capítol  $\hat{r}$  prenen valors d'aquestes unitats que els fan més semblants als anteriors al capítol  $\hat{r}$  que als posteriors. Per tant hi ha la possibilitat que en la part final del llibre hi hagi barreja d'estils. En aquest capítol es resumeixen els resultats obtinguts en els capítols anteriors, en les anàlisis realitzades per capítols de més de 200 paraules.

### 14.1 Punt de canvi per capítols

La taula 14.1 conté un resum de les estimacions del punt de canvi obtingudes en l'anàlisi de les seqüències formades pels valors que prenen les unitats d'estadística textual en els capítols de més de 200 paraules.

Observem com, per la majoria de les unitats considerades, les estimacions del punt de canvi, tant per les seqüències multinomials com per algunes seqüències univariants, es troben entre el capítol 371 i el 382. L'excepció la trobem en l'ús de lletres, on les quatre estimacions no coincideixen, i en la llargada de frase, on l'estimació balla entre els capítols 263 i 269.

Per estimar el punt de canvi fem servir el gràfic de la mesura de bondat de l'ajust en funció de  $r$  (el màxim del logaritme de la versemblança per seqüències binomials i multinomials i l'estadístic  $F$  de la taula ANOVA per seqüències Normals). Permet visualitzar el màxim que correspon a l'estimador  $\hat{r}$  del punt de canvi. Observem com

per algunes de les unitats hi ha un màxim local de valor proper al màxim global. A la taula 14.1 escrivim el punt pel que obtenim el màxim local entre parèntesi.

<b>Unitat</b>		<b>Model</b>	<b>Punt de canvi</b>
<i>Llargada de paraula</i>	<i>Llargada mitjana</i>	<i>Normal</i>	345(371)
	<i>Ús paraules 1 lletra</i>	<i>Logístic</i>	387
	<i>Ús paraules 2 lletres</i>	<i>Logístic</i>	327
	<i>Ús paraules 9 lletres</i>	<i>Logístic</i>	371
	<i>Ús paraules 10 o més lletres</i>	<i>Logístic</i>	345
	<i>Distribució Llargada</i>	<i>Politòmic</i>	371(345)
<i>Llargada de frase</i>	<i>Llargada mitjana</i>	<i>Normal</i>	263
	<i>Distribució Llargada</i>	<i>Politòmic</i>	269
<i>Ús de Lletres</i>	<i>24 Lletres</i>	<i>Politòmic</i>	300
	<i>5 Vocals</i>	<i>Politòmic</i>	371
	<i>36 Lletres</i>	<i>Politòmic</i>	299 (369)
	<i>14 Vocals</i>	<i>Politòmic</i>	368(383)
<i>Ús de Paraules</i>	<i>Ús d'<u>e</u> (més freqüent)</i>	<i>Logístic</i>	382
	<i>Ús de <u>de</u> (2<sup>a</sup> més freqüent)</i>	<i>Logístic</i>	381
	<i>Ús de <u>la</u> (3<sup>a</sup> més freqüent)</i>	<i>Logístic</i>	356
	<i>12 més freqüents</i>	<i>Politòmic</i>	382
	<i>25 més freqüents</i>	<i>Politòmic</i>	382
<i>Diversitat</i>	<i>Índex Simpson, D</i>	<i>Normal</i>	382
	<i>Nombre de tipus, V</i>	<i>Normal</i>	382
	<i>Entropia, H<sup>s</sup></i>	<i>Normal</i>	382

Taula 14.1: Quadre resum amb l'estimació del punt de canvi en les seqüències de dades més representatives del *Tirant*. En les dues primeres columnes hi tenim les unitats analitzades, en la tercera el model ajustat per a l'estimació. En la darrera hi tenim l'estimació del punt del canvi, i en parèntesi hi ha el punt de la seqüència pel que obtenim un màxim local amb valor proper al global.

## 14.2 Assignació de capítols a un dels dos estils

Encara que es detecti la presència d'una frontera d'estil, sempre és possible que alguns dels capítols, o blocs, quedin mal classificats. En els capítols anteriors hem realitzat anàlisi cluster per cadascuna de les unitats analitzades. A l'apartat 14.2.1 es resumeixen aquests resultats, al 14.2.2 es refan les anàlisis cluster imposant la restricció que en totes les anàlisis la grandària dels dos clusters siguin idèntiques, i a l'apartat 14.2.3 es fa una única anàlisi cluster a partir de mesures resum de les unitats d'estadística textual emprades, fent servir l'algorisme k-means per agrupar els capítols en dos clusters.

### 14.2.1 Resum dels resultats de les anàlisi clusters individuals

L'Annex 14.1 conté les assignacions dels capítols a una de les dues agrupacions obtingudes de les set anàlisi cluster realitzades per a les unitats:

- llargada de paraula, en el capítol 8,
- ús de 36 lletres i de 5 vocals, en el capítol 11,
- ús de les 12 i 25 paraules més freqüents i de les paraules que hem observat que tenen proporcions significativament diferents entre abans i després del capítol 382 del *Tirant* (i que anomenem discriminants), en el capítol 12,
- diversitat de vocabulari.

Per a la darrera, els resultats són els obtinguts fent servir l'algorisme *k-means* ponderant les observacions, i en les altres s'ha emprat la tècnica cluster per a les files d'una taula de contingència basada en la distància  $\chi^2$ .

En l'Annex 14.1 codifiquem, per cadascuna de les set anàlisis realitzades, amb *0* els capítols assignats al cluster que conté majoritàriament els capítols del principi del llibre, i amb *1* els assignats al cluster que conté majoritàriament els capítols del final i que anomenem cluster *del principi* i *del final*, respectivament. La columna etiquetada com **S7** conté la suma de les set columnes anteriors, de manera que un capítol que té *0* en aquesta columna ha estat assignat en les set anàlisis al cluster *del principi*, un capítol que té *1* en aquesta columna ha estat assignat en sis de les set anàlisis al cluster *del principi* i en una al *del final*, mentre que un capítol que té *7* en aquesta columna ha estat assignat en les set anàlisis al cluster *del final*. La columna **S3** correspon a la suma dels resultats de l'anàlisi cluster per llargada de paraula, paraules discriminants i diversitat i la columna **S5** a la suma de **S3** amb els resultats obtinguts per l'ús de les 12 i de les 25 paraules més abundants.

El gràfic de la figura 14.1 conté l'evolució del valor que pren la columna **S7** al llarg del *Tirant*. Observem com a partir del capítol 383, la majoria han estat assignats en cinc, sis o set anàlisis al cluster *del final*. Els capítols 412, 432 i 435 són els únics posteriors 382 que han sigut assignats en els set casos al cluster *del principi*. Dels capítols anteriors al 382, només el 52 ha sigut agrupat en les set anàlisis en el cluster *del final*.

En l'evolució al llarg del *Tirant* de la columna **S5** observem com pel que fa als capítols posteriors al 382 que han sigut assignats en les cinc anàlisis al cluster *del principi*, cal afegir als anteriors el 411, 433 i 475, i dels capítols anteriors al 382 classificats en el cluster *del final* cal afegir només el 44. En tots quatre casos, l'anàlisi que ha donat agrupaments diferents ha sigut el de les 5 vocals.

Pel que fa a l'evolució al llarg del *Tirant* del valor de la columna **S3**, no hem d'afegir cap capítol a la llista dels posteriors al 382 que han sigut assignats en les tres anàlisis al cluster *del principi*. Pel que fa als capítols anteriors al 382 que han sigut assignats en les tres anàlisis en el cluster *del final* cal afegir el 7 i el 194.

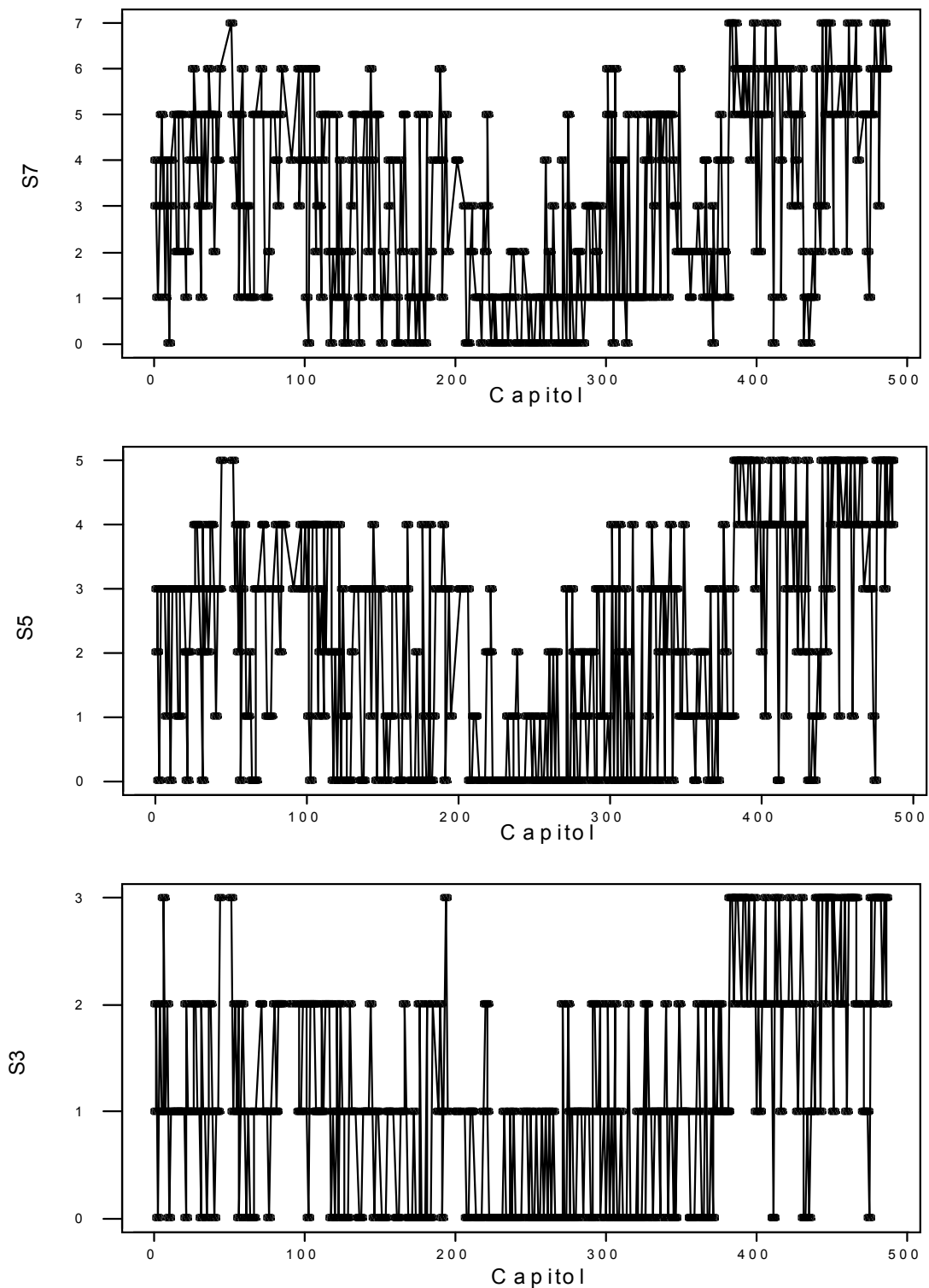


Figura 14.1 Evolució del valor de les columnes *S7* (a dalt), *S5* (al mig) i *S3* (a baix) de l'Annex 14.1., i que contenen la suma de les columnes corresponents a les anàlisis cluster realitzats en els capítols anteriors per la llargada de paraula, per paraules discriminants, i per diversitat, *S3*, aquestes tres més les columnes corresponents a l'ús de 12 i 25 paraules eina per *S5*, i la suma de le set columnes per *S7*.

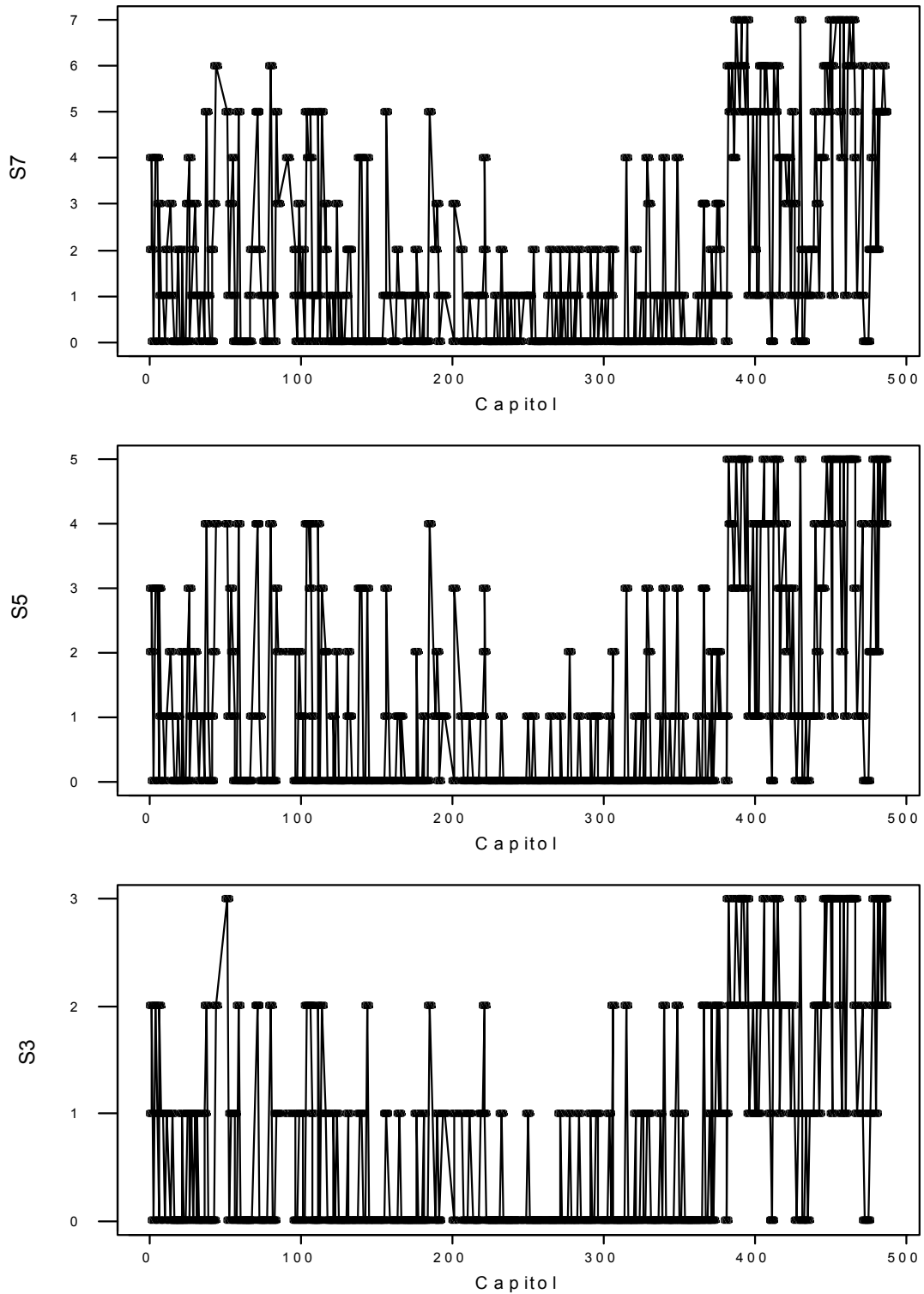


Figura 14.2. Evolució del valor de les columnes *S7* (a dalt), *S5* (al mig) i *S3* (a baix) de l'Annex 14.1, i que contenen la suma de les columnes corresponents a les anàlisis cluster realitzades imposant que la grandària del cluster *del final* sigui 100, per la llargada de paraula, per paraules discriminants, i per diversitat, *S3*, aquestes tres més les columnes corresponents a l'ús de 12 i 25 paraules eina per *S5*, i la suma de les set columnes per *S7*.

## 14.2.2 Anàlisi clusters individuals imposant grandària 100

A l'estudi realitzat en l'apartat anterior observem com el nombre de capítols agrupats en el cluster que conté majoritàriament els capítols del final del *Tirant* és molt variable, anant dels 101 capítols per la diversitat del vocabulari als 236 per l'ús de 5 vocals. Per tal de ser restrictius i identificar aquells capítols que estem més segurs que tenen trets estilístics més diferents dels "del principi" del llibre, s'ha realitzat una anàlisi en la que imposem que el nombre de capítols assignats al cluster *del final* sigui 100. Si anomenem *del principi* al cluster que conté majoritàriament els capítols del principi del llibre i *del final* al que conté majoritàriament els capítols del final del llibre, l'algorisme és el que es detalla en el capítol 7 de la tesi amb petites modificacions:

- L'assignació inicial es fa de manera que els primers 382 capítols són assignats al cluster *del principi* i els capítols del 383 al final al cluster *del final*, d'acord amb els resultats obtinguts de l'estimació del punt de canvi.
- En primer lloc es busca si hi ha algun capítol assignat al cluster *del final* que canviant d'assignació faci augmentar la distància  $\chi^2$  entre els dos clusters. Si n'hi ha més d'un, canvia d'assignació el que provoca un augment més gran en la distància  $\chi^2$ . Quan no n'hi ha cap, es busca el capítol assignat al cluster *del principi* que canviant d'assignació provoqui l'augment màxim en la distància  $\chi^2$  entre els dos clusters.
- L'algorisme para quan no hi ha canvis d'assignació que provoquin un augment en la distància  $\chi^2$  entre els dos clusters o be quan el nombre de capítols assignats al cluster *del final* és 100.

S'ha triat 100 com a grandària màxima del cluster *del final* perquè és un valor molt proper tant al nombre de capítols posteriors al punt de canvi i perquè és el mínim de les grandàries obtingudes de les anàlisis realitzades a 14.2.1, i que correspon a l'anàlisi cluster per les mesures de diversitat.

Els gràfics de la figura 14.2 i l'Annex 14.2 són els anàlegs pels resultats d'aquesta anàlisi als de la figura 14.1 i l'Annex 14.1, respectivament. En els gràfics hi trobem l'evolució al llarg del *Tirant* del valor de la suma de les columnes que hem etiquetat en l'Annex 14.2 com **S3**, **S5** i **S7** i que contenen, respectivament, el resultat de l'anàlisi cluster per la llargada de paraula, ús de paraules discriminats i diversitat a baix, aquests tres més ús de les 12 i les 25 paraules més freqüents al mig, i d'aquests cinc més l'ús de 36 lletres i 5 vocals a dalt. Recordem que en aquestes columnes, un 0 senyala que el capítol ha sigut agrupat en el cluster *del principi* del llibre i un 1 al que conté majoritàriament els capítols del final del llibre.

El gràfic de dalt de la figura 14.2 conté l'evolució al llarg del *Tirant* de la columna etiquetada com **S7**. hi observem com a partir del 383, la majoria de capítols han estat assignats en cinc, sis o set de les anàlisis al cluster *del final*. Els capítols 411, 412, 428, 432, 433, 473, 474 i 475 són els únics posteriors 382 que han sigut assignats en els set casos al cluster *del principi*. Els capítols 383, 385, 388, 391, 392, 393, 395, 404, 405, 406, 408, 413, 415, 430, 447, 448, 450, 452, 454, 456, 459, 462, 463, 464, 465, 471, 479 i 485 són els posteriors 382 que han sigut assignats en sis o set casos (subratllats) al cluster *del final*. Dels capítols anteriors al 382, cap ha sigut agrupat en les set anàlisis en el cluster *del final*, i els ho han estat en sis de les set anàlisis.

En el gràfic del mig de la figura 14.2, que conté l'evolució al llarg del *Tirant* de la columna etiquetada com **S5**, observem com, pel que fa als capítols posteriors al 382 que



han sigut assignats en les cinc anàlisis al cluster *del principi*, cal afegir als anteriors el 435 i a la llista d'agrupats en les cinc anàlisis al cluster *del final* afegim el 466, 483 i 487, i cap dels capítols anteriors al 382 ha sigut classificats en el cluster *del final* en tots cinc anàlisis.

La columna **S3** correspon a la suma dels resultats obtinguts per la llargada de paraula, l'ús de les paraules discriminants, i la diversitat de vocabulari. En el gràfic de la figura 14.2 observem com no hem d'afegir cap capítol als assignats en les tres anàlisis al cluster *del principi* i com cal afegir el capítol 466 als assignats en les tres anàlisis al cluster *del final*. Pel que fa als capítols anteriors al 382 que han sigut assignats en les tres anàlisis en el cluster *del final* hi trobem, només, el 52.

### 14.2.3 Cluster amb llargada de paraula, ús de paraules i diversitat

Hem fet una anàlisi cluster dels capítols del Tirant mitjançant l'algorisme k-means per variables contínues i donant als capítols un pes proporcional a la seva llargada. Hem pres com a variables l'índex de Simpson i la primera component de l'anàlisi de correspondències per la llargada de paraula i per les paraules discriminants. D'aquesta manera les tres unitats que han senyalat el punt de canvi estan representades per una variable i tenen, per tant, el mateix pes en l'anàlisi.

L'Annex 14.3 conté en quin cluster han estat agrupats els capítols, codificant amb *0* el cluster que conté majoritàriament els capítols del principi i amb *1* el del final del *Tirant*. El gràfic de la figura 14.3 mostra l'assignació de capítols a clusters.

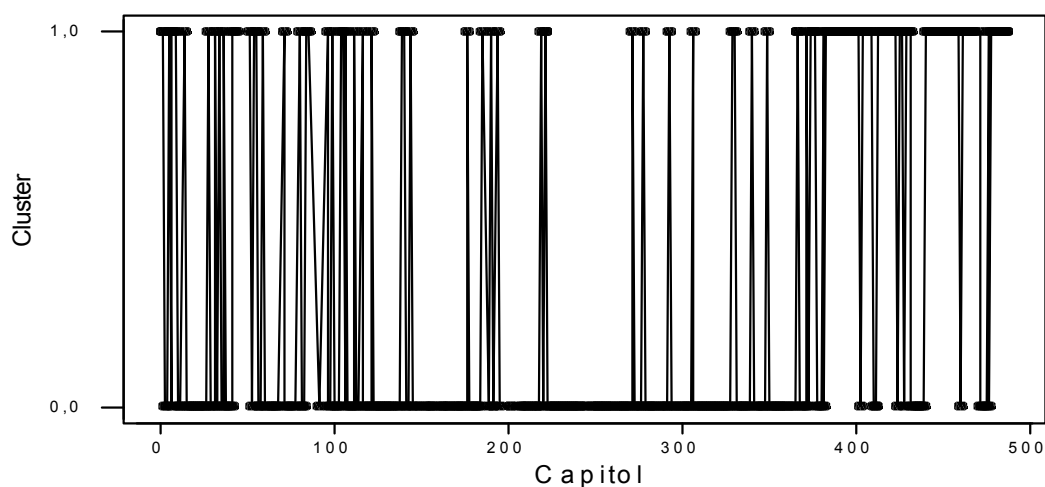


Figura 14.3. Assignació dels capítols a un dels dos clusters, després de l'anàlisi cluster basada en l'algorisme *k-means* ponderat pels capítols de més de 200 paraules del *Tirant*.

Hi ha vint capítols posteriors al 382 que han sigut agrupats en el cluster *del principi*. Són el 403, 410, 411, 412, 424, 428, 432, 433, 434, 435, 436, 437, 438, 439, 460, 472, 473, 474, 475 i 477. Observem com hi ha tres seqüències 410-412, 432-439 i 472-477 que han estat assignats al cluster *del principi*. Aquest resultat coincideix força amb el obtingut en els apartats anteriors.

Pel que fa als capítols anteriors al 382, n'hi ha 56 agrupats en el cluster *del final*, dels quals sis entre els nou primers, sis més entre els capítols 41 a 47 i cinc més entre els capítols 372 i 377. Recordem que per la llargada de paraula l'estimació del punt de canvi és el capítol 371.

### 14.3 Conclusions

Per tot el que hem anat trobant estem convençuts que hi ha un canvi en l'estil entre els capítols 371 i 382, que difícilment pot ser atribuïble a l'argument. Aquest canvi, a més a més, és sobtat, i per tant no pot ser degut a una evolució de l'estil al llarg del període en que va ser escrit el *Tirant*.

També hem trobat que, després del punt canvi d'estil, conviuen capítols amb els dos estils, el que probablement reforça la teoria de el segon autor va afegir capítols sobre un original pràcticament acabat. A la taula 14.2 hi trobem el llistat dels capítols que estan mal classificats pel punt de canvi. En el llistat hem sigut restrictius, i només hi consten aquells pels que hi ha evidències molt sòlides.

Anteriors al 382	Posteriors al 382
44, 52, 80	403, 411, 412, 424, 432, 433, 434, 435, 436, 437, 438, 439, 460, 472, 473, 474, 475

Taula 14.2: Capítols mal classificats pel punt de canvi. En el llistat hi trobem els anteriors amb estil més semblant al del final del llibre, i els posteriors amb estil més proper al del principi.

Per atacar el problema, hem començat adaptant tècniques d'anàlisi descriptiva de dades, com els gràfics de control i l'anàlisi de correspondències. Per explotar la base de dades, proposem un mètode pràctic per estimar un o més d'un punt de canvi en seqüències de normals, de binomials i de multinomials. El mètode es basa en l'ajust de models i troba els estimadors màxim versemblants del(s) punt(s) de canvi. També hem utilitzat un mètode cluster basat en l'ajust de models per a dades politòmiques, per a agrupar les files d'una taula de contingència.

Pel que fa a les unitats lexicomètriques, la llargada de paraula, l'ús de paraules freqüents i lliures del context ens han sigut molt útils per a l'estimació del punt de canvi i l'atribució d'estil als capítols. Abordem l'anàlisi de la riquesa i diversitat de vocabulari mitjançant l'estudi comparatiu de 12 maneres diferents de mesurar-les. L'ús de lletres, tot i ser menys útil, serveix per a reforçar l'evidència del que trobem amb les unitats abans esmentades. La llargada de frase no ens ha sigut útil, com tampoc no ho havia sigut en estudis estilomètrics anteriors, com els de Mosteller, F. i Wallace, D.L. (1964) sobre els *Federalist Papers*. La llargada de capítol, tot i el que diuen filòlegs com Coromines (1956), no ha sigut útil per a determinar una frontera d'estil en el *Tirant*.