NEW TENDENCIES IN GEOGRAPHICAL DIALECTOLOGY: THE CATALAN *CORPUS ORAL DIALECTAL* (COD)*

ESTEVE CLUA & MARIA-ROSA LLORET
Universitat Pompeu Fabra & Universitat de Barcelona

1. *Introduction*

Traditional dialect grouping was done on the basis of qualitative criteria, usually related to the concept of 'isogloss' or 'bundles of isoglosses'. As known, a major flaw of this method is that it takes into account a very limited number of words or linguistic features, and the decision on which words and features are selected is rather controversial. More recently, with the use of data computerization and statistical techniques, a much bigger percentage of items can be taken into account, and several methods related to data quantification have been developed for dialect grouping. The most important difference of the quantitative approach with respect to the qualitative one is that the statistical method does not assign a priority qualitative ranking of the variables used for classificatory purposes; rather, all variables *weight* alike, they are all equally measured. Therefore, the quantitative approach lies on the global measurement of the variables found in a big set of data, and not on a small selection of variables and data. A well-known drawback of this method though is that, since each of the items used for measurement is equal to each of the others, the measure does not indicate structural similarity between varieties but it just counts superficial coincidences or differences. A way of correcting this defect would be to weight the items in a different way, but it is claimed that such mechanism would introduce arbitrariness in the methodology and, for this reason, quantitative studies continue leaving aside structural differences. (Cf. Francis 1983.)

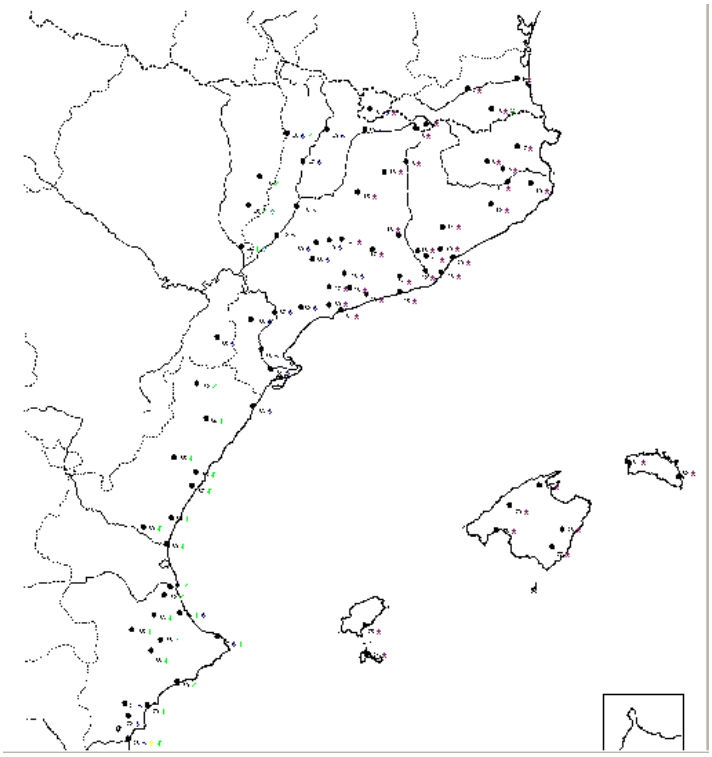
The goal of this paper is to show that there is a principled way of capturing structural (i.e., qualitative) differences within a quantitative approach and that the result of applying such methodology to dialect grouping is more accurate than the one obtained on purely qualitative or quantitative grounds. The paper is organized as follows. First, we will present the characteristics of the corpus on which our study is based, the Catalan *Corpus Oral Dialectal*. Second, we will introduce the main tenets of the quantitative dialectometrical methodology. Third, we will show the kind of linguistic analysis we pursue to distinguish regular phonetic facts from underlying differences and we will evaluate the consequences of such distinction for dialectometrics. Finally, we will illustrate the results of our research by reviewing the dialectal classification of Valencian Catalan.

* This research is sponsored by the Spanish Ministerio de Educación y Ciencia and the FEDER (research project HUM2004-01504/FILO: "Explotación de un corpus oral dialectal (ECOD)"). It also benefits from funding by the Catalan Government (Departament d'Universitats, Recerca i Societat; Generalitat de Catalunya; research group 2001SGR0004). More information about this project is available online at: http://www.ub.edu/lincat.

2. *The Catalan* Corpus Oral Dialectal *(COD)*

Throughout the last fifteen years the Departament de Filologia Catalana of the Universitat de Barcelona has gathered and systematized in databases a corpus of contemporary Catalan --the *Corpus Oral Dialectal* (COD). Data were collected with computerization in mind through a questionnaire of approximately 600 phonetic and morphological items and recordings of 10-minute samples of casual speech. The fieldwork was carried out in each of the 86 county towns of the whole Catalan-speaking area, throughout Spain, Andorra, the southeast part of France and the city of Alghero in Sardinia, Italy (see Figure 1).

Figure 1: *General map*



We interviewed 2-3 speakers in each town. The use of a third speaker was designed in order to be able to select the majority form in cases with variation where data computerization required a single answer. The informants were 30-45 years old, middle class speakers, with a minimum amount of formal education.

The selection of localities and speakers was done with the purpose of recording the common mode of speaking of the inhabitants of more urban areas, where the population concentrates nowadays. Our aim, thus, is different from that of traditional surveys, which concentrate on recording old speakers of small rural areas that preserve the indigenous varieties of a language that are usually in danger of being lost due to the pressure of education, media and the standard variety, among others.

The results of the questionnaire have been systematized in databases, which nowadays contain 135.480 phonetic items and 532.508 morphological items. Up to now, 50 free-speech samples have been orthographically and phonetically transcribed and aligned with their corresponding sound files (cf. Viaplana & Perea 2003). (More details on the characteristics of the corpus appear in Lloret & Perea 2002.)

From the corpus, we have developed the following three different lines of research:

a) We are making the questionnaire data accessible through computerized maps (cf. Perea 2005).

b) We analyze the questionnaire data from the phonological and morphological views, and the free-speech data from the syntactic view as well (cf., among others, Bonet & Lloret 2005a, b; Grimalt 2002; Lloret 2004; Pons 2004a, b; Querol 2004). We also develop studies related to language variation and linguistic change by comparing our data with those in old questionnaires and atlases (cf., among others, Campmany 2004a, b; Lloret 2003).

c) We use the analyzed data from the questionnaire to develop dialect-grouping techniques based on a multivariate analysis, in line with dialectometrics and the cluster analysis (cf. Clua 1999a, b, 2005; Viaplana 1999). As said, in this paper, we are going to focus on this last issue.

3. ***Dialectometric approach to dialect grouping***

The crucial notion of the quantitative approaches to dialect grouping is the concept of 'linguistic distance', which is the measurement of the set of similarities between dialects; or, conversely, the measurement of dissimilarities or distance. The similarities or dissimilarities taken into account are the variables of the statistical analysis. (Cf., among others, Séguy 1973 and Goebl 1992.)

The methodology used to calculate the linguistic distance is the following. First, one has to establish comparative matrixes that relate the set of linguistic variables analyzed with the set of localities taken into account (Table1).

Table 1. *Comparative matrix*

Variables / Localities	X_1	X_2	X_3	...	X_p
Locality 1	X_{11}	X_{12}	X_{13}	...	X_{1p}
Locality 2	X_{21}	X_{22}	X_{23}	...	X_{2p}
Locality 3	X_{31}	X_{32}	X_{33}	...	X_{3p}
...	...	...	...	...	...
Locality n	X_{n1}	X_{n2}	X_{n3}	...	X_{np}

From these matrixes, the coincidences between varieties are calculated in order to get the similarity matrixes. Before that, though, we need to establish the similarity index from which the computer program can calculate the coincidences and automatically determine the linguistic distance between localities. The choice of this index is very important for the results, since they may vary according to the kind of index used. This is especially relevant when

the comparative matrixes contain null cells (due to the lack of information) or when they contain multiple answers. Traditionally, this problem was difficult to overcome because the data used for the analysis were not homogeneous because they were not collected with computerization in mind; rather dialectometrical studies usually take as point of departure data from the linguistic atlases or from old questionnaires. In our project this problem has been completely overcome due to the fact that data were collected with computerization in mind from the beginning, and thus there are no blanks and a single majority form per item can be selected when this is required for the computer program.

Generally the index of similarity that is used regards the percentage of coincidences with respect to the total number of elements compared among two varieties. Under certain conditions, the index can be as simple as assigning the value '1' if two localities *i* and *j* coincide with respect to a variable *k*; and '0' otherwise (see (1)).

(1) Measure of similarity

$$s(i,j) = \sum_k coin_k(i,j)$$

Where $coin_k$ (i,j) takes value 1 when, regarding the linguistic variable *k*, the localities *i* and *j* coincide, and it takes value 0 otherwise.

The similarity matrixes calculate the number of coincidences between localities with respect to the set of variables taken into account. They measure the linguistic distance between varieties. An example of this kind of matrixes is provided in Table 2 (see the complete matrix in the appendix).

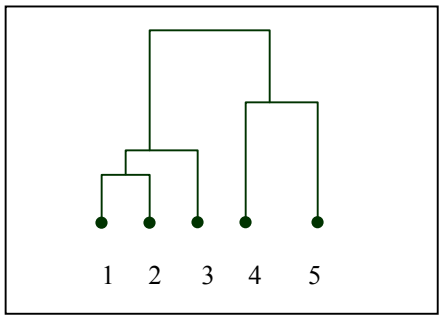
Table 2: *Example of similarity matrixes*

	Guardamar	Elx	Novelda	Alacant	la Vila	Alcoi	Cocentaina	Ontinyent
Guardamar	0,00							
Elx	2123	0,00						
Novelda	2124	2233,17	0,00					
Alacant	2050	2156,42	2189,56	0,00				
la Vila	2054	2093,53	2149,57	2130,54	0,00			
Alcoi	2028	2028,13	2102,14	2090,47	2235,47	0,00		
Cocentaina	2053	2033	2107	2140	2256	2236	0,00	
Ontinyent	2020	2051	2115	2140	2268	2284	2258	0,00

These similarity matrixes allow figuring out easily the localities that share a high number of coincidences, that is, the ones that are linguistically closer; or, conversely, the ones that are linguistically farther. However, except for comparison between two varieties, in the previous representation it is difficult to foresee the overall interpretation of the results or to get a plausible, easy

visualization of the linguistic distances between all varieties. For this reason, we have to resort to another kind of representation, which is able to represent in a single level (as in a plane) the original structure of the linguistic distances (which are multidimensional) with a minimal distortion of the data. One of the more usual techniques applied in such cases is the 'cluster analysis', which allows an optimal representation of the grouping results. The purpose of this kind of representation is to build clusters based on a measurement of similarities that is fixed by an algorithm. This allows a tree representation, or 'dendrogram', as the one shown in Figure 2.

Figure 2: *Example of dendrogram*



The main element of this kind of analysis is the algorithm used to convert the numerical data into a cluster representation. The one we use is the well-known UPGMA (Unweighted Pair-Group Method using Arithmetic Averages; cf. Sneath & Sokal 1973).

4. ***Linguistic analysis and linguistic distance between dialects***

Traditionally, the quantitative approach as well as the qualitative one is item centered and superficially oriented, that is, it is based on the phonetic outputs. Computer technologies have broadened the linguistic research on dialectology in the sense that it is now possible to deal with enormous amounts of data, but in order to exploit them qualitatively as well we need a mechanism to introduce structural differences. The way we propose to add the qualitative countenance to the quantitative approach is to have a means to weight, in addition to the surface differences, the real underlying differences that are due to the effects of the regular phenomena that a language display. We will next illustrate this point by looking at the shape of the second person singular pronominal clitic in Valencian Catalan (2). (The clitic appears in bold in the examples below.)

(2)		Variety 1	Variety 2	Variety 3	
a.	*et parle*	[**et**párle]	[**te**párle]	[**te**párle]	"I talk to you"
b.	*t'anime*	[**t**aníme]	[**t**aníme]	[**t**aníme]	"I cheer you up"
c.	*animant-te*	[animán**te**]	[animán**te**]	[animán**te**]	"cheering you up"
d.	*anima't*	[anima**t**]	[anima**t**]	[anima**te**]	"cheer you up!"

The examples in (2) show that all the varieties display a non-syllabic form [t] and different syllabic forms: one with a vowel before the consonant ([et]) and another one with the vowel after the consonant ([te]). In variety 1, [et]

appears before a verb that begins with a consonant (2a), while [te] appears after a verb that ends in a consonant (2c). In variety 2, [te] shows up in these two cases (2a, c). In variety 3, [te] further appears after a verb ending in vowel (2a, c, d). From the point of view of traditional approaches, the linguistic distance between these three varieties is very similar. All the varieties show the same forms in (2b) and (2c). Varieties 1 and 3 differ in two cases: (2a) and (2d). Variety 2 differs in one form with respect to variety 1, (2a), and in one other form with respect to variety 3, (2d).

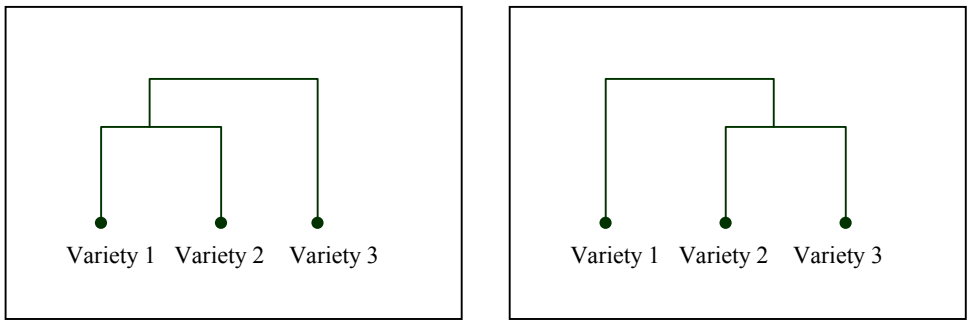
From the point of view of surface approaches, the distance between varieties 1 and 2 has the value 1 because they differ in one form only: [etpárle] vs. [tepárle]. Varieties 1 and 3 show a linguistic distance of 2 because they differ in two forms: [etpárle] vs. [tepárle] and [anímat] vs. [anímate]. Varieties 2 and 3 display a value of 1 because they differ in one form: [anímat] vs. [anímate]. (See the similarity matrix in Table 3.)

Table 3: *Similarity matrix based on phonetic data*

	Variety 1	Variety 2	Variety 3
Variety 1			
Variety 2	1		
Variety 3	2	1	

According to these data, in a dendrogram (or tree representation), either varieties 1 and 2 are grouped closer than 3, because they differ in 1 form only, or varieties 2 and 3 are clustered closer than 1, because they also differ in 1 form only (Figure 3).

Figure 3: *Dendrogram based on phonetic data*



All this variation, though, can be reduced on the basis of syllabification by distinguishing underlying differences from the ones that are due to regular phenomena. Under this view, varieties 1 and 2 have a single underlying form (/t/) and the vowel [e] is inserted in order to satisfy syllabic requirements. That is, as in other contexts, epenthesis applies when the addition of the clitic to the verb creates a sequence that cannot be properly syllabified. The difference between these two dialects lies in the position of the epenthesis. In variety 1, the epenthetic vowel always appears at the periphery of verb-clitic sequences, i.e. at the beginning in (3a) but at the end in (3c) In variety 2 instead it is always placed to the right of the clitic, i.e. [te] in (3a) and (3c). Variety 3 is

completely different. The crucial example here is the last one, i.e. *anima-te* [anímate] (3d). In this case, the verb ends in a vowel and, thus, there is no syllabic reason to assume that the vowel of the clitic is inserted through epenthesis to repair syllabification. For this variety, it is more coherent to establish that the underlying form of the clitic contains the vowel (/te/), although this vowel deletes when it appears in contact with another vowel, cf. *t'anime* [taníme] in (3b). This vowel also deletes in other vocalic contexts in the language (cf. *entr**e** **a**mics*: *entr[a]mics* "between friends", ***no** **é**s tan gran*: *n[o]s tan gran* "it is not that big").

(3)		Variety 1 /t/	Variety 2 /t/	Variety 3 /te/	
a.	*et parle*	[**et**párle]	[**te**párle]	[**te**párle]	"I talk to you"
b.	*t'anime*	[**t**aníme]	[**t**aníme]	[**t**aníme]	"I cheer you up"
c.	animant-te	[animán**te**]	[animán**te**]	[animán**te**]	"cheering you up"
d.	anima't	[anima**t**]	[anima**t**]	[anima**te**]	"cheer you up!"

In other words, variety 3 has preserved the old shape of the clitic (/te/, from the Latin form *te*), but in certain contexts the vowel deletes in accordance with the regular phonology of the language. Unlike variety 3, varieties 1 and 2 have re-structured their system. They show a single-consonant underlying form (/t/) that undergoes epenthesis for syllabic reasons. Therefore, the linguistic distance between varieties 1 and 2 is indeed smaller than that with variety 3, which has a different underlying representation. We will next show how our analysis captures this fact.

The similarity matrix presented in Table 4 shows that, as for underlying differences concerning the four forms under study, varieties 1 and 2 have zero differences (both have a /t/ underlying form), but variety 1 with respect to 3, and 2 with respect to 3 show 4 differences (variety 3 departs from a /te/ underlying form).

Table 4: *Similarity matrix based on the phonological analysis (I): Underlying differences:* /t/$_{1, 2}$ *vs.* /te/$_3$

	Variety 1	Variety 2	Variety 3
Variety 1			
Variety 2	0		
Variety 3	4	4	

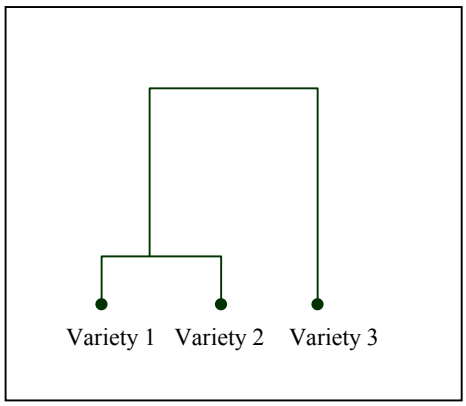
The similarity matrix presented in Table 5 further calculates the differences concerning the phenomena involved. Here, varieties 1 and 2 differ only in the position of the epenthesis. Varieties 1 and 3, and varieties 2 and 3 differ in displaying or not epenthesis and vowel deletion.

Table 5: *Similarity matrix based on the phonological analysis (II). Differences in the phenomena involved*

	Variety 1	Variety 2	Variety 3
Variety 1			
Variety 2	1		
Variety 3	2	2	

In accordance with our analysis, the resulting dendrogram (Figure 4) shows a closer relation between varieties 1 and 2, and, significantly, a larger distance between these two and variety 3.

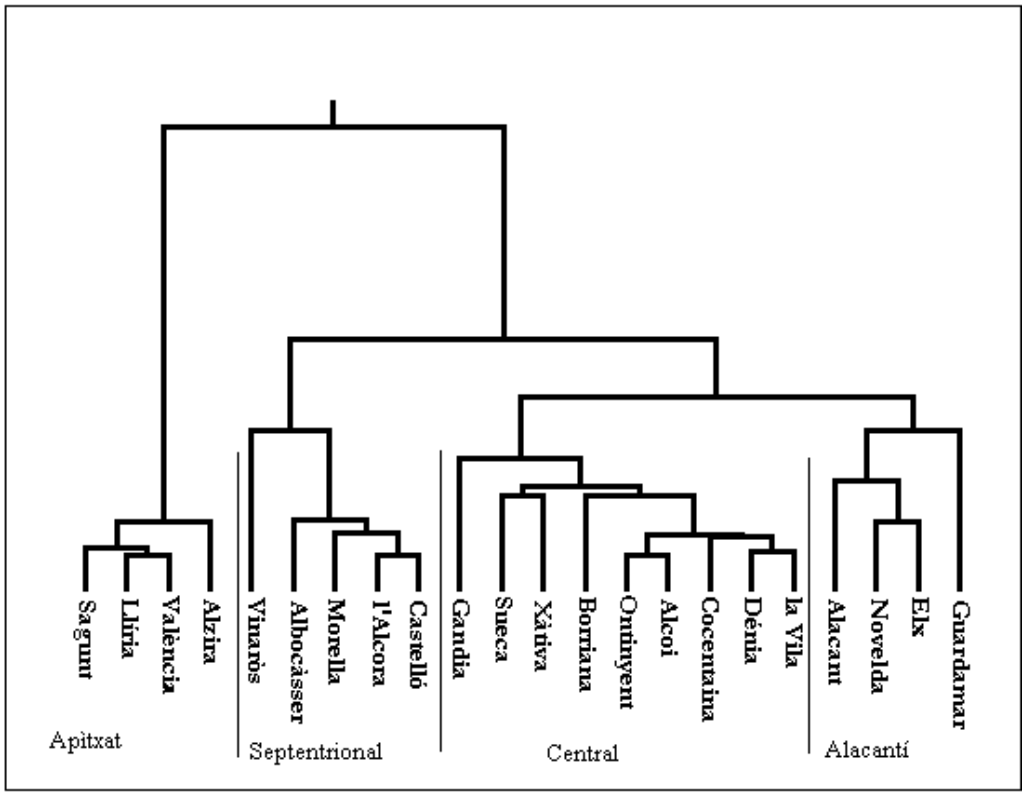
Figure 4: *Dendrogram based on the phonological analysis*



5. ***Example: The case of Valencian Catalan***

We will next illustrate the results of applying this methodology to a whole set of data. The example is taken from Clua (1999a, b), where he analyzes the inflection of Valencian Catalan based on the data of our corpus. Clua uses a dialectometric approach to review the traditional dialectal classification (which is based on the notion of bundles of isoglosses), but he applies dialectometrics to the already analyzed data of our corpus, along the lines we have previously illustrated. The overall results of this study are shown in the following dendrometrical representation (Figure 5), extracted from the similarity matrix presented in the appendix (excerpted in Table 2):

Figure 5: *Dendrometrical representation for Valencian based on inflectional data*



In Figure 5, it is clear that the variety on the left side (*apitxat*) shows the biggest linguistic distance with respect to the other three. As for the other big group, the two varieties that appear on the right side (*alacantí* and *central*) are closer than the other one (*septentrional*) is. Downwards, we end up having four groups: *valencià apitxat* ("Tight Valencian", which is the traditional term to refer to the varieties that show sibilant devoicing); *valencià septentrional* ("Northern Valencian"); *valencià central* ("Central Valencian"), and *valencià meridional* or *alacantí* ("Southern Valencian").

In the traditional classifications of Catalan four groups are distinguished too (cf. Colomina 1999). However, while in the traditional approach, which is based on the cartography of certain isoglosses, all four groups are considered to be at the same linguistic distance, in our dendrometric representation (cf. Figure 5) the grouping is much more accurate. In addition to that, a closer look at the results shows significant differences with respect to the scope of each dialect. For the sake of comparison, we present these two classifications in map

form: The map in Figure 6 shows the geographical distribution of our results and the one in Figure 7 that of traditional classifications.

Figure 6: *Map of the dendrometric classification*

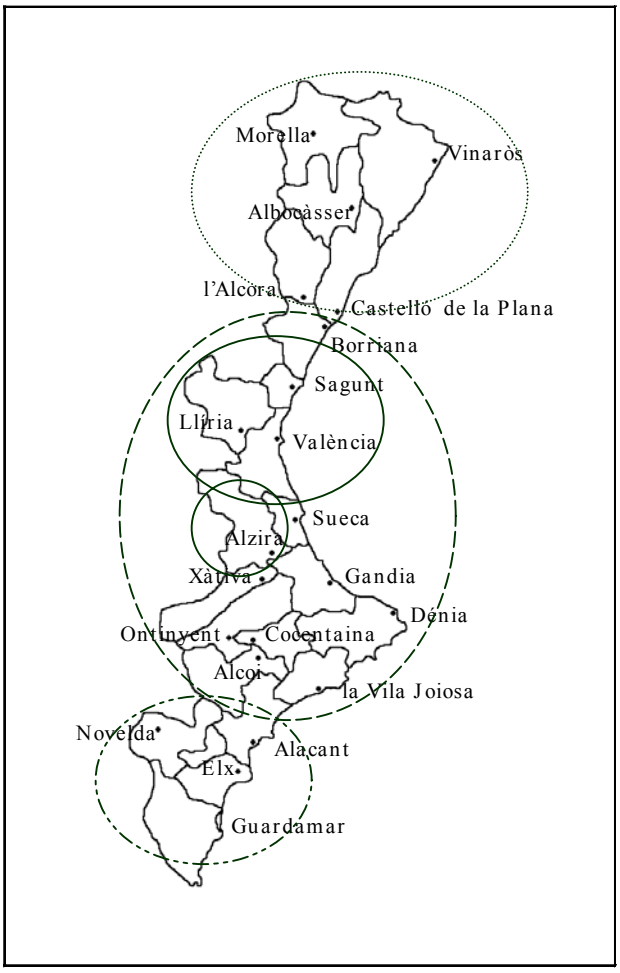
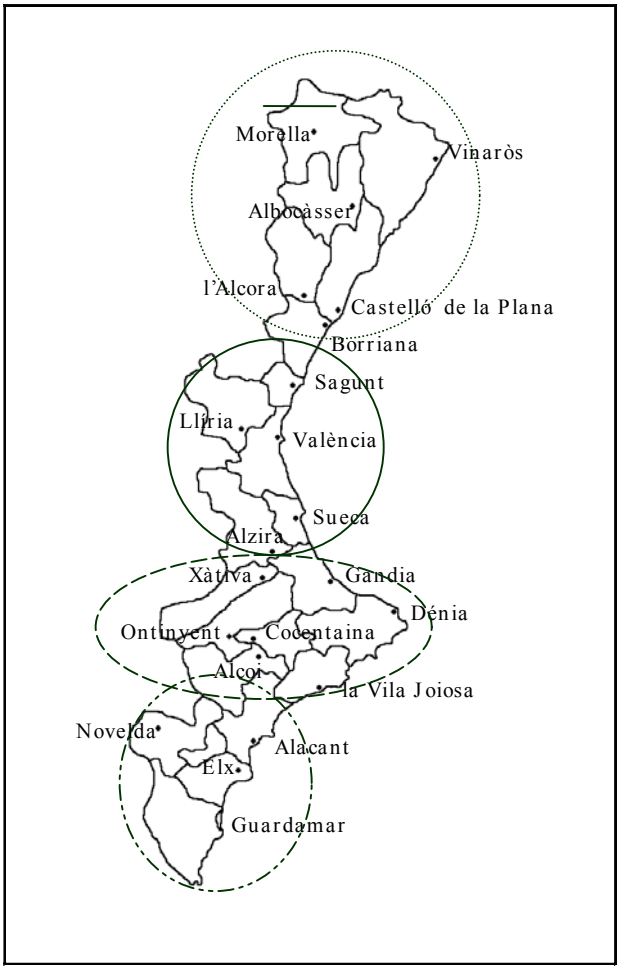


Figure 7: *Map of the traditional classification*



By comparing the two maps, note that Southern Valencian (the one containing the city of *Alacant*) remains almost alike in both approaches. However, the geographic scope of the other dialects is quite different. Northern Valencian (the one containing the city of *Castelló de la Plana*) goes further South in the traditional approach. Also, the area of Tight Valencian (the one containing the city of *València*) is much bigger in the traditional classification, while in ours the area of Central Valencian (the one containing the city of *Gandia*) is bigger, running from North to South of the central area.

6. ***Conclusion***

In sum, we believe that the dialectal grouping made on the bases of dialectometry with previously analyzed data gives us a better picture of the linguistic distance between dialects. It also allows us to weight the crucial discriminatory facts of each system, which, in the absence of such distinction, remain amalgamated in a simple sum of distinct surface forms.

APPENDIX

	Guardamar	Elx	Novelda	Alacant	la Vila	Alcoi	Cocentaina	Ontinyent	Dénia	Xàtiva	Gandia	Sueca	Alzira	València	Llíria	Sagunt	Borriana	Castelló	l'Alcora	Albocàsser	Morella	Vinaròs
Guardamar	0,00																					
Elx	2123,63	0,00																				
Novelda	2124,10	2233,17	0,00																			
Alacant	2050,98	2156,42	2189,56	0,00																		
la Vila	2054,43	2093,53	2149,57	2130,54	0,00																	
Alcoi	2028,40	2028,13	2102,14	2090,47	2235,47	0,00																
Cocentaina	2053,43	2033,73	2107,97	2140,78	2256,34	2236,63	0,00															
Ontinyent	2020,37	2051,89	2115,59	2140,46	2268,29	2284,44	2258,36	0,00														
Dénia	2035,42	2057,06	2119,67	2174,67	2277,76	2245,34	2258,11	2273,11	0,00													
Xàtiva	1957,67	2045,40	2027,28	2074,29	2180,75	2180,77	2148,92	2199,83	2210,17	0,00												
Gandia	1907,30	1947,83	1974,04	1983,39	2120,83	2137,18	2108,69	2146,91	2152,28	2181,46	0,00											
Sueca	1982,85	1983,54	2022,05	2016,68	2161,50	2194,71	2160,45	2201,93	2202,27	2195,23	2166,90	0,00										
Alzira	1541,23	1559,29	1592,42	1650,25	1711,54	1717,51	1701,67	1743,88	1718,26	1802,06	1874,66	1787,42	0,00									
València	1583,51	1576,13	1579,38	1650,45	1711,06	1710,04	1696,98	1742,98	1739,82	1769,82	1867,48	1811,18	2213,25	0,00								
Llíria	1606,23	1566,07	1544,64	1623,47	1677,43	1677,47	1663,34	1710,55	1695,82	1781,39	1828,81	1787,05	2253,89	2284,18	0,00							
Sagunt	1594,60	1562,59	1551,68	1629,73	1683,80	1690,54	1717,07	1722,74	1702,18	1787,32	1833,31	1794,95	2235,11	2260,84	2283,56	0,00						
Borriana	1942,07	1989,00	2023,17	2095,58	2163,69	2213,31	2166,30	2229,84	2205,12	2175,48	2100,14	2151,55	1739,00	1731,29	1698,11	1711,63	0,00					
Castelló	1809,33	1812,33	1856,43	1938,55	2005,73	2056,57	1994,37	2035,50	2037,30	2044,10	1939,28	1995,90	1529,90	1546,15	1496,90	1503,93	2086,19	0,00				
l'Alcora	1788,51	1806,56	1871,10	1877,28	1987,89	2043,88	1984,53	2021,90	2011,37	2014,60	1918,66	1971,47	1524,83	1523,14	1483,05	1495,63	2102,19	2283,74	0,00			
Albocàsser	1820,02	1773,10	1824,03	1827,98	1969,91	2008,59	2006,57	2004,70	2003,90	2002,57	1906,53	1980,72	1523,71	1530,83	1489,71	1530,35	2051,95	2232,68	2228,60	0,00		
Morella	1772,80	1823,14	1849,98	1862,69	1978,47	2009,70	1949,66	2001,59	2017,14	2005,50	1906,95	1957,67	1515,27	1521,06	1481,60	1491,46	2085,27	2245,63	2257,03	2227,97	0,00	
Vinaròs	1898,97	1857,53	1900,08	1992,75	2035,80	2084,78	2073,76	2061,15	2087,25	2043,45	1974,59	2021,05	1600,33	1604,12	1566,67	1603,37	2153,70	2091,55	2090,73	2110,41	2097,26	0,00

REFERENCES

Bonet, Eulàlia & Maria-Rosa Lloret. 2005a. "Against serial evaluation in Optimality Theory". *Lingua* 115. 1303-1323.

---------. 2005b. "More on Alignment as an alternative to domains: The syllabification of Catalan clitics". *Probus* 17:1. 37-78. (Prepublication version available online at: http://rucs.rutgers.edu/roa.html, ROA No. 592)

Campmany, Elisenda. 2004a. *Estudi de la variació dels clítics pronominals en textos formals i informals del català central septentrional*. MA thesis, Universitat de Barcelona.

---------. 2004b. "Nova aproximació a l'estudi dels proclítics del català central septentrional". Paper presented at the 50th Annual Conference of the Anglo-Catalan Society / 11th Colloquium of the North American Catalan Society, Eton College, December 2004.

Clua, Esteve. 1999a. *Variació i distància lingüística. Classificació dialectal del valencià a partir de la morfologia flexiva*. Ph.D. dissertation, Universitat de Barcelona..

---------. 1999b. "Distància lingüística i classificació de varietats dialectals". *Caplletra* 26. 11-26.

---------. 2005. "El mètode dialectomètric: aplicació de l'anàlisi multivariant a la classificació de les varietats del català". *Dialectologia i recursos informàtics* ed. by Maria Pilar Perea, 59-87. Barcelona: PPU. [With a CD-ROM.]

Colomina, Jordi. 1999. *Dialectologia catalana. Introducció i guia bibliogràfic*a. Alacant: Universitat d'Alacant.

Francis, W.N. 1983. *Dialectology. An Introduction*. London & New York: Longman.

Goebl, Hans. 1992. "Problèmes et méthodes de la dialectométrie actuelle (avec application à l'AIS)". *Nazioarteko Dialektologia Biltzarra Euskaltzaindia. Bilbo 1991* 10. 21-25.

Grimalt, Pere. 2002. *Enclisi i accentuació en el català de Mallorca i Menorca. Descripció i anàlisi*. MA thesis, Universitat de Barcelona.

Lloret, Maria-Rosa. 2003. "Variació condicionada: Aspectes qualitatius i quantitatius". *Catalan Review* XVIII:2.

---------. 2004. "The phonological role of paradigms: The case of insular Catalan". *Contemporary Approaches to Romance Linguistics. Selected papers from the 33rd Linguistic Symposium on Romance Languages (LSRL), Bloomington, Indiana, April 2003* ed. by Julie Auger, J. Clancy Clements & Barbara Vance, 275-297. Amsterdam & Philadelphia: John Benjamins. (Prepublication version available online at: http://rucs.rutgers.edu/roa.html, ROA No. 646)

--------- & Maria Pilar Perea. 2002. "A Report on the *Corpus Oral Dialectal del Català Actual* (COD)". *Dialectologia et Geolinguistica* 10. 1-18.

Perea, Maria Pilar. 2005. "Mapes electrònics i mapes sonors". *Dialectologia i recursos informàtics* ed. by Maria Pilar Perea, 135-152. Barcelona: PPU. [With a CD-ROM.]

Pons, Clàudia. 2004a. "Segmental and featural strategies to avoid adjacent sibilant segments in Balearic Catalan: An Optimality-theoretic account".

Romance Languages and Linguistic Theory. Selected papers from 'Going Romance', Groningen, November 2002 ed. by Reineke Bok-Bennema, Bart Hollebrandse & Brigitte Kampers-Manhe, 217-234. Amsterdam & Philadelphia: John Benjamins.

---------. 2004b. *Els contactes consonàntics en balear. Descripció i anàlisi*. Ph.D. dissertation, Universitat de Barcelona.

Querol, Laia. 2004. *Distribució i valor morfemàtic dels segments velars en la morfologia verbal del català occidental*. MA thesis, Universitat de Barcelona.

Séguy, J. 1973. "La dialectometrie dans l'Atlas linguistique de la Gascogne". *Revue de linguistique romane* 37. 1-24.

Sneath, P. H. A. & R. R. Sokal. 1973. *Numerical Taxonomy. The Principles and Practice of Numerical Classification. A Series of Books in Biology*. San Francisco: W. H. Freeman and Company.

Viaplana, Joaquim. 1999. *Entre la dialectologia i la lingüística*. Barcelona: Publicacions de l'Abadia de Montserrat.

--------- & Maria Pilar Perea, eds. 2003. *Textos orals dialectals del català sincronitzats. Una selecció*. Barcelona: PPU. [With a CD-ROM.]